

Kemiallisen kommunikaation geneettinen tausta ja kemosensoorigeenien evoluutio muurahaisissa



Katri Ketola

Pro gradu

Perinnöllisyystieteen osasto

Biotieteiden laitos

Helsingin yliopisto

Huhtikuu 2015

Tiedekunta – Fakultet – Faculty Bio- ja ympäristötieteellinen tiedekunta		Laitos – Institution– Department Biotieteiden laitos
Tekijä – Författare – Author Katri Ketola		
Työn nimi – Arbetets titel – Title Kemiallisen kommunikaation geneettinen tausta ja kemosensoirigeenien evoluutio muurahaisissa		
Oppiaine – Läroämne – Subject Perinnöllisyystiede		
Työn laji – Arbetets art – Level Pro gradu	Aika – Datum – Month and year Huhtikuu 2015	Sivumäärä – Sidoantal – Number of pages 59
<p>Tiivistelmä – Referat – Abstract</p> <p>Darwinistista luonnonvalintaa voidaan tutkia geenisekvensseistä. Positiivinen luonnonvalinta suosii tiettyä geenimuotoa, joka siten yleistyy populaatiossa. Vahvan positiivisen luonnonvalinnan havaitseminen on harvinaista, mutta sen on todettu vaikuttavan erityisesti immuunipuolustuksen ja aistihavainnoinnin geeneihin. Positiivisen valinnan alaisilla hajugeeneillä saattaa olla rooli lajiutumisessa ja sopeutumisessa. Tämän vuoksi hajuaistimukseen perustuva kemiallinen kommunikaatio on kiinnostava kohde luonnonvalinnan mittaamiselle. Sosiaaliset hyönteiset, kuten muurahaiset, ovat kemiallisen kommunikaation malliorganismeja. Ne käyttävät kemiallista kommunikaatiota paitsi ravinnon hankintaan ja saalistajien tunnistamiseen, myös yhdyskunnan toimintojen organisoimiseen tuhansien yksilöiden välillä.</p> <p>Muurahaiset ja muut hyönteiset aistivat hajumolekyyliä tuntosarvillaan. OBP:t (odorant binding protein) ja CSP:t (chemosensory protein) toimivat hajumolekyylien sitomisessa ja niiden kuljettamisessa tuntosarvien aistinelinten hemolymfassa. OBP- ja CSP-proteiinit saattavat toimia myös hajumolekyylien valikoimisessa. Tässä työssä tutkittiin luonnonvalintaa neljässä geenissä, jotka ovat konservoituneita muurahaislajien välillä. Kolmen niistä (OBP1, CSP1, CSP7) tiedetään ilmentyvän vahvasti muurahaisten tuntosarvissa, mikä tukee oletusta niiden roolista hajuviestinnässä. Lisäksi CSP7-geenillä on tunnettu tehtävä pesätoverien tunnistuksessa muurahaisilla ja OBP1-geenillä kuningatarferomonin sitomisessa mehiläisellä. Aineisto sisälsi geenisekvenssejä seitsemästä <i>Formica</i>-muurahaislajista, yhteensä n. 270 yksilöstä.</p> <p>Työn tavoitteina oli selvittää 1) muuntelun määrää näissä geeneissä läheisten muurahaislajien välillä ja lajien sisällä, 2) mitkä evoluutiovoimat, luonnonvalinta vai satunnaisajautuminen, ovat muuntelun takana ja 3) onko kahden sosiaalisen muodon (monogyniset ja polygyniset populaatiot) välillä systemaattisia eroja OBP- ja CSP-geeneissä.</p> <p>Aineistolle suoritettiin sekä muuntelun määrää kuvaavia analyysejä että evoluutioanalyysejä. Muuntelua lajien välillä kaikissa neljässä geenissä visualisoitiin fylogeneettisellä puulla, pääkoordinaattianalyyseillä ja lajien välisillä fiksoituneilla eroilla. Lisäksi lajien sisäisiä populaatioiden välisiä eroja tutkittiin parittaisilla F_{ST}-arvoilla. Tietoa populaatioiden eroavuudesta käytettiin evoluutioanalyyseihin. Evoluutiovoimien analysointi tehtiin McDonald-Kreitman, Tajiman D, Fu ja Li sekä MFDM -testejä käyttäen.</p> <p>Lajien välistä muuntelua kuvaavat analyytit näyttävät kahden muurahaislajin, <i>F. exsecta</i> ja <i>F. cinerea</i>, eroavan selvästi <i>F. rufa</i> -ryhmän lajeista. Sen sijaan viisi <i>F. rufa</i> -ryhmän lajia eivät merkitsevästi eroa toisistaan tämän aineiston perusteella. Tämä voi johtua siitä, että lajiutumisesta on vain vähän aikaa tai lajien välisistä hybrideistä aineistossa. Luonnonvalinnan testien tulokset eivät ole yhdenmukaisia. CSP7-geenissä on kuitenkin eniten merkkejä mahdollisesta valinnasta. Erityisesti yhdessä ennustetussa transkriptiofaktorin sitoutumiskohdassa havaittiin mahdollista valintaa ja sekvenssimuuntelua <i>F. cinerea</i>ssa. Valinnan voimat saattavat siis vaikuttaa CSP7-geenin säätelyyn. Jatkotutkimuksissa olisi kiinnostavaa tarkistaa todelliset transkriptiofaktoreiden sitoutumiskohdat.</p>		
Avainsanat – Nyckelord – Keywords Kemiallinen kommunikaatio, kemosensoirigeeni, CSP, OBP, hajuaisti, evoluutio, valinta, adaptaatio, geenisäätely, sosiaalisuus		
Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors Jonna Kulmuni		
Säilytyspaikka – Förvaringställe – Where deposited		
Muita tietoja – Övriga uppgifter – Additional information		

SISÄLLYSLUETTELO

1. LYHENTEET	5
2. JOHDANTO	7
2.1. Luonnonvalinta ja sen mittaaminen sekvenssiaineistosta	7
2.2. Hyönteisten hajugeenit ja hajuaistimus	8
2.3. Tutkimuksessa käytettävien hajugeenien funktiot	13
2.4. OBP-geenien yhteys muurahaisten sosiaaliseen muotoon	14
3. TYÖN TAVOITTEET	15
4. MATERIAALIT JA MENETELMÄT	16
4.1. Sekvenssiaineisto	16
4.2. Käytetyt menetelmät	17
4.2.1. Geenien rinnastus ja annotointi	17
4.2.2. Sekvenssimuuntelua ja lajien välisiä eroja kuvaavat analyysit	18
4.2.3. Evoluutioanalyysit	20
5. TULOKSET	25
5.1. Geenien rakenne ja lokus	25
5.2. Sekvenssimuuntelu lajeissa	27
5.2.1. Fiksoituneet muutokset lajien välillä	27
5.2.2. Fylogenia	28
5.2.3. Pääkoordinaattianalyysi	29
5.2.4. Nukleotididiversiteetti	30
5.3. Sekvenssimuuntelu populaatioissa	31
5.3.1. F_{ST}	31
5.3.2. Monogyniset ja polygyniset populaatiot	32
5.4. Luonnonvalinta	33

5.4.1. CSP1	33
5.4.2. CSP7	35
5.4.3. OBP1	38
5.4.4. OBP10	40
5.5. Rekombinaatio	42
6. TULOSTEN TARKASTELU.....	43
6.1. Sekvenssimuuntelu <i>F. rufa</i> -ryhmän lajien välillä on vähäistä	43
6.2. Luonnonvalinta hajugeeneissä	44
6.2.1 CSP7-geenissä on eniten merkkejä luonnonvalinnasta	45
6.2.2. Geenisäätelyyn kohdistuva luonnonvalinta	48
6.3. Muurahaisten sosiaalisten muotojen välinen muuntelu	48
6.4. Käytettyjen menetelmien soveltuvuus	49
6.5. Johtopäätökset ja jatkotutkimukset	50
 7. KIITOKSET	 51
8. LÄHTEET.....	52
9. LIITTEET.....	59

1. LYHENTEET

AIC	HyPHY-ohjelmiston rekombinaatiotestikriteeri pienelle aineistolle
AmelCSP1	Mehiläisen (<i>A. mellifera</i>) CSP-proteiini, joka vastaa muurahaisen CSP7-proteiinia
AmelCSP4	Mehiläisen (<i>A. mellifera</i>) CSP-proteiini, joka vastaa muurahaisen CSP1-proteiinia
ASP1	Mehiläisen kuningatarferomonia sitova OBP-geeni (antennal-specific protein 1)
B-faktori	<i>Drosophila</i> -kärpäsen transkriptiofaktori
BIC	HyPHY-ohjelmiston rekombinaatiotestikriteeri suurelle aineistolle
CHC	Kutikulaarinen hiilivety (cuticular hydrocarbon)
CSP	Kemosensoriproteiini (chemosensory protein)
Dfd	<i>Drosophila</i> -kärpäsen transkriptiofaktori (Deformed)
DnaSP	Ohjelmisto nukleotidipolymorfismin analysointiin rinnastetusta DNA-sekvenssiaineistosta (DNA Sequence Polymorphism)
DSXF	<i>Drosophila</i> -kärpäsen transkriptiofaktori (Doublesex female)
DSXM	<i>Drosophila</i> -kärpäsen transkriptiofaktori (Doublesex male)
Eve	<i>Drosophila</i> -kärpäsen transkriptiofaktori (Even-skipped)
F _{ST}	Erilaistumista mittaava indeksi, osapopulaation odotettu heterotsygotia verrattuna koko populaation odotettuun heterotsygotiaan (fixation index)
Ftz	<i>Drosophila</i> -kärpäsen transkriptiofaktori (Fushi tarazu)
Gp-9	Tulimuurahaisen OBP-proteiini, jolla on yhteys muurahaisten sosiaaliin muotoihin (General protein-9)
Hb	<i>Drosophila</i> -kärpäsen transkriptiofaktori (Hunchback)
HyPhy	Geneettisten sekvenssien analysointiohjelmisto
Mad	<i>Drosophila</i> -kärpäsen transkriptiofaktori (Mothers against dpp)
MAFFT	Usean sekvenssin rinnastusohjelma (Multiple Alignment using Fast Fourier Transform)
MEGA	Ohjelma joka suorittaa geneettisiä molekyyli evoluutioanalyysyjä (Molecular Evolutionary Genetics Analysis)
MFDM	Uusien mutaatioiden maksimifrekvenssiä käyttävä evoluutiotesti (Maximum Frequency of Derived Mutations)

MK	McDonald-Kreitman -evoluutiotesti
NI	McDonald-Kreitman -testiin liittyvä neutraalisuusindeksi (neutrality index)
OBP	Hajumolekyyliä sitova proteiini (odorant binding protein)
OR	Hajuaistinreseptori (odorant receptor)
PCoA	Pääkoordinaattianalyysi, menetelmä aineiston yksilöiden tai ryhmien samankaltaisuuden visualisointiin koordinaatistossa (Principal Coordinate Analysis)
PHASE	Ohjelma joka jakaa diploidin sekvenssiaineiston alleeleiksi
Prd	<i>Drosophila</i> -kärpäsen transkriptiofaktori (Paired)
PROMO	Transkriptiofaktoreiden sitoutumiskohtia ennustava ohjelma
SBP	HyPhy-ohjelmiston rekombinaatiotesti (Single Break Point)
SNP	Yhden nukleotidin polymorfismi (Single Nucleotide Polymorphism)
TII	<i>Drosophila</i> -kärpäsen transkriptiofaktori (Tailless)
TRANSFAC	Transkriptiofaktoreiden sitoutumiskohtien sekvenssitietokanta
UTR	Lähetti-RNA:n transloimattomat 5' ja 3'-sekvenssit (untranslated region)
Zen-1	<i>Drosophila</i> -kärpäsen transkriptiofaktori (Zerknullt)

2. JOHDANTO

2.1. Luonnonvalinta ja sen mittaaminen sekvenssiaineistosta

Darwinin esittämän luonnonvalinnan periaatteen mukaisesti ympäristöön parhaiten sopeutuvat yksilöt lisääntyvät ja siirtävät geenejään seuraaviin sukupolviin. Sopeutumisen taustalla ovat perinnölliset muutokset geeneissä tai muualla perimässä, joten luonnonvalintaa voidaan mitata nukleotidisekvensseistä. Uudet mutaatiot synnyttävät jatkuvasti geneettistä muuntelua yksilöiden välillä. Neutraalin (Kimura 1968) ja sen korvanneen lähes neutraalin teorian (Ohta 1973; Ohta 1992) mukaan suurin osa uusista mutaatioista on neutraaleja, eivätkä ne vaikuta yksilön kelpoisuuteen (fitness). Tällöin sekvenssimuuntelu syntyy uusien mutaatioiden ja geneettisen satunnaisajautumisen tasapainosta. Neutraalien alleelien poistumiseen tai fiksoitumiseen populaatiossa vaikuttaa vain sattuma eli geneettinen satunnaisajautuminen. Lähes neutraali teoria hyväksyy heikon valinnan. Sen mukaan tiettyjen alleelien yleistyminen populaatiossa voi johtua heikosta puhdistavasta valinnasta lievästi haitallisia alleeleja vastaan. Lähes neutraali teoria huomioi myös sen, että satunnaisajautumisella on suurempi vaikutus pienissä populaatioissa. Neutraalista ja lähes neutraalista teoriasta on muodostunut molekyyli evoluution nollahypoteesi, joka täytyy kumota, jotta luonnonvalintaa voidaan todeta.

Luonnonvalinta voi olla positiivista, puhdistavaa tai tasapainottavaa (Hartl ym. 2007). Kahdessa ensimmäisessä tapauksessa populaatiossa havaitaan odotettua vähemmän muuntelua. Positiivinen valinta ajaa hyödyllisen alleelin nopeaa fiksoitumista ja puhdistava valinta haitallisten alleelien poistumista. Molemmissa tapauksissa tuloksena on neutraaliin teoriaan verrattuna odotettua pienempi lukumäärä alleeleja. Positiivinen valinta edistää sopeutumista yleistämällä selviytymistä ja lisääntymistä parantavia alleeleja populaatiossa. Positiivinen valinta voi ilmetä myös geenin monistumisena. Tasapainottava (balansoiva) valinta taas ylläpitää muuntelua ja polymorfismit siis säilyvät pidempään kuin odotettaisiin neutraalin teorian mukaan. Puhdistavaa valintaa havaitaan näistä yleisimmin. Noin 13,5 % ihmisen geeneistä on havaittu heikkoa puhdistavaa valintaa tai tasapainottavaa valintaa (Bustamante ym. 2005). Noin 0,9 % mehiläisen geeneistä on vahvan puhdistavan valinnan alaisena (Harpur ym. 2014). Suurimman osan (noin 90 %) jäniksen uusista aminohappoa muuttavista mutaatioista arvioidaan olevan haitallisia ja puhdistavan valinnan alaisena (Carneiro ym. 2012). Lukuarvojen ristiriitaisuus johtuu eroista käytetyissä tutkimusmenetelmissä sekä siitä, tarkastellaanko koko genomia, kaikkia geenejä vai pelkästään ei-synonyymisiä mutaatiota. Käsitykset positiivisen valinnan määrästä ovat vaihdelleet

ja asiaa on voitu tutkia genomilaajuisesti vasta viime aikoina, kun aineistoa on alkanut kertyä useammilta lajeilta.

Luonnonvalinnan mittaamiseen DNA-sekvenssistä on kehitetty erilaisia testejä, jotka perustuvat yleensä joko alleelifrekvenssispektriin tai aminohappoa muuttavien (ei-synonyymisten) ja muuttamattomien (synonyymisten) mutaatioiden suhteeseen. Synonyymisten muutosten oletetaan olevan valinnan kannalta neutraaleja, koska ne eivät muuta proteiinia tai sen rakennetta. Jos ei-synonyymisiä mutaatioita on huomattavasti synonyymisiä mutaatioita enemmän suhteessa ei-synonyymisten nukleotidikohtien lukumäärään, mutaatioiden täytyy olla hyödyllisiä ja luonnonvalinta nopeuttaa nukleotidikorvautumisia. Tässä tapauksessa siis todetaan positiivista luonnonvalintaa.

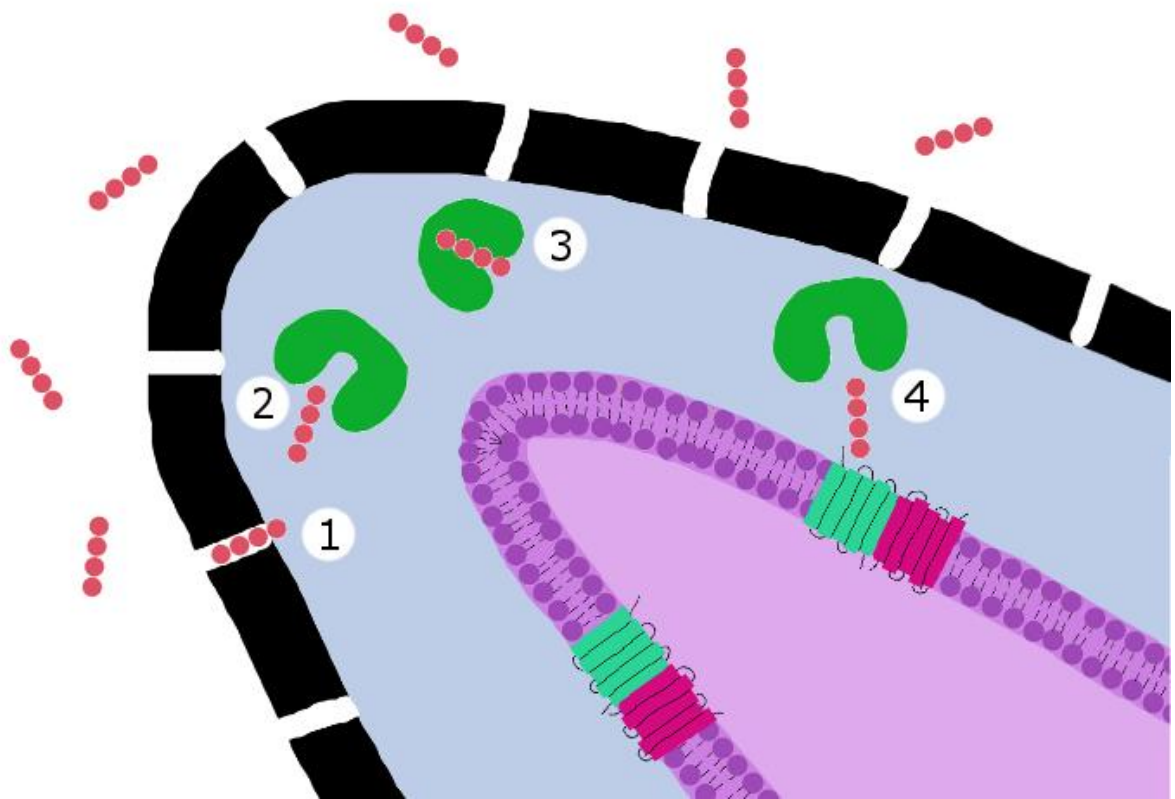
Perinteiset valinnan testit eivät yleensä ole tehokkaita ja tarvitsevat suuren aineiston, jotta neutraali nollahypoteesi voidaan kumota. Neutraalin teorian hylkäämiseen voi olla luonnonvalinnan lisäksi muita syitä: heterogeenisyys mutaatioprosessissa, virheet näytteiden otossa, populaation piilorakenne ja migraatio (Hartl ym. 2007: s. 318). Luonnonvalinta ja populaation epävakaa demografia, kuten vaihteleva populaatiokoko tai sukupuolijakauma, jättävät genomiin samankaltaisia merkkejä (Li 2011, Li ym. 2012) ja niiden erottaminen toisistaan evoluutiotestien tuloksista on vaikeaa. Perinteisesti on ajateltu, että demografian vaikutus on sama koko genomien alueella ja että luonnonvalinta vaikuttaa genomisekvenssiin vain paikallisesti. Tämän periaatteen mukaan valinta voidaan erottaa demografiasta vertaamalla tutkittavaa geeniä suuren alueeseen tai koko genomiin. Yksittäisten geenien evoluutiota tutkittaessa tästä ei ole apua. Lisäksi menetelmä ei toimi, jos genomissa onkin paljon valintaa. Esimerkiksi *Drosophila*-kärpäspopulaatioilla, joilla on suuri efektiivinen populaatiokoko, on todettu erityisen tehokasta positiivista valintaa (Andolfatto 2005). Kun populaatiokoko on suuri, geneettinen hitchhiking eli hyödyllisen mutaation kanssa kytkeytyneiden neutraalien alueiden fiksoituminen, voi vaikuttaa laajasti ja se voidaan tulkita väärin demografian vaikutukseksi (Li ym. 2012). Myös rekombinaatio voi vaikeuttaa valinnan tunnistamista (Hartl ym. 2007: s. 239).

Vahvan positiivisen luonnonvalinnan havaitseminen on yleensä harvinaista, mutta sen on todettu vaikuttavan tiettyihin geeniryhmiin. Hetkittäinen positiivinen valinta voi olla yleistä, mutta osa valinnan testeistä havaitsee vain harvinaisempaa jatkuvaa positiivista valintaa. Mehiläisellä on tarkasteltu viimeaikaista valintaa koko genomista ja noin yhdeksän prosenttia geeneistä on positiivisen valinnan alaisena (Harpur ym. 2014). Yhdeksässä prosentissa ihmisen geeneistä on havaittu positiivista valintaa (Bustamante ym. 2005). Noin 45 % *Drosophila*-kärpäsen (Smith ym.

2002) ja 60 % jäniksen (Carneiro ym. 2012) aminohappoa muuttaneista nukleotidikorvautumisista arvioidaan fiksoituneen positiivisen valinnan ajamana. Positiivista valintaa on todettu muita geenejä yleisemmin esimerkiksi immuunipuolustukseen (Salazar-Jaramillo ym. 2014, Nei 2005, Ferreira ym. 2007) ja aistihavainnointiin (Whiteman ym. 2008, Forêt ym. 2006, Vieira ym. 2012, Kulmuni ym. 2013a, Kulmuni ym. 2013b) liittyvissä geeneissä. Immuunigeenit ovat jatkuvan valinnan alaisina isännän ja parasiitin koevoluution vuoksi: parasiitit koittavat päästä immuunipuolustuksen läpi ja isäntä puolestaan estää sen. Hajugeenien puolestaan oletetaan muuttuvan lajiutumisen yhteydessä, koska lajiutumisessa yleensä siirrytään uuteen ympäristöön tai erikoistutaan uuteen ravinnonlähteeseen (Smadja ym. 2009). Tämän vuoksi hajuaistimukseen perustuva kemiallinen kommunikaatio on kiinnostava alue luonnonvalinnan ja erityisesti positiivisen valinnan mittaamiselle. Kemiallisen kommunikaation malliorganismeina sosiaaliset hyönteiset, kuten muurahaiset, ovat luonteva tutkimuskohde. Sosiaaliset hyönteiset tarvitsevat kommunikaatiota esimerkiksi yhdyskunnan toiminnan ohjaamiseen, ravinnon löytämiseen ja saalistajien, pesätoverien ja oman pesän yksilöiden kastien tunnistamiseen. Sosiaalisten hyönteisten yleisin kommunikaation muoto on tehokas ja nopea kemiallinen kommunikaatio (Richard ym. 2013), joka perustuu hajugeeniperheisiin.

2.2. Hyönteisten hajugeenit ja hajuaistimus

Hyönteiset aistivat hajumolekyyliä tuntosarvillaan. Hyönteisten tuottamat feromonit ja muut ilmassa liikkuvat hajumolekyylit kulkeutuvat tuntosarvien karvamaisten aistinelinten hemolymfaan (vaihe 1 Kuvassa 1). Lymfassa on tunnistettu kaksi pienten vesiliukoisten proteiinien luokkaa: hajua sitovat proteiinit (OBP, odorant binding protein) ja kemosensoriproteiinit (CSP, chemosensory protein). Ne liuottavat ja sitovat hajuaistinelinten huokosista sisään tulevat hajumolekyylit sitoutumiskohtaansa ja auttavat hydrofobisten molekyylien kuljetuksessa vesiympäristössä (vaihe 2 Kuvassa 1). Proteiinit kuljettavat hajumolekyylit neuronin tuojahaaraketta ympäröivän onkalon aistinlymfan läpi neuronimembraanin hajuaistinreseptoreille (OR, odorant receptor) (vaihe 3 Kuvassa 1). Joko hajumolekyyli itsessään tai hajumolekyylin ja kuljettajaproteiinin kompleksi (Laughlin ym. 2008) aktivoi neuronin reseptorin selektiivisesti (vaihe 4 Kuvassa 1) ja viesti välittyy eteenpäin hermoimpulssina. Aistinlymfaan jäävät hajumolekyylit inaktivoidaan.



Kuva 1. Kaavakuva OBP- ja CSP-proteiinien roolista hyönteisten hajuaistimuksessa. 1) Hajumolekyylit (kuvassa punaisia) kulkeutuvat hajuaistinlymfaan hajuaistinelinten kutikulan (kuvassa musta) huokosista. 2) Hajumolekyylit sitoutuvat OBP- tai CSP-proteiineihin (kuvassa vihreitä), jotka 3) kuljettavat hajumolekyylit aistinlymfan vesifaasin läpi neuronin reseptorille (kuvassa heteromeerinen reseptorikompleksi koostuu vihreästä ja punaisesta osasta). 4) Hajumolekyyli vapautuu kuljettajaproteiinista ja aktivoi reseptorin. (Kuvan malli: Leal 2013)

OBP-proteiinien ligandin sitominen ja vapauttaminen johtuvat pH-riippuvaisesta konformaatio-muutoksesta. Hajuaistinreseptorien läheisyydessä on alhainen pH, jonka aiheuttavat alueelle kerääntyvät protonit, jotka neutraloivat dendriitin pinnan negatiivista varausta (Leal 2013). OBP-proteiinin rakenteeseen kuuluu C-terminaalinen häntä, joka muodostaa α -heliksin alhaisessa, lähes neutraalissa pH:ssa ja sitoutuu proteiinin sitoutumistaskuun estäen ligandin sitoutumisen (Xu ym. 2008). Korkeassa pH:ssa α -heliksirakenne purkautuu ja sitoutumistasku on vapaana ligandille. OBP- ja CSP-proteiinien tehtävä kuljettajina ja hajuaistimuksen herkkyydessä on tunnettu (Swarup ym. 2011), mutta niiden rooli hajumolekyylien valikoimisessa on sen sijaan epävarma.

Hyönteisten ja selkärankaisten OBP:t toimivat molemmat kuljetuksessa, mutta eivät ole samaa fylogeneettistä alkuperää. Nisäkkäillä on suuri määrä hajureseptoreita (noin 1000) (Prasad ym. 1999) ja vain muutama OBP-geeni (Tegoni ym. 2000). Nisäkkäillä hajumolekyylien selektiivistä

vastaanottoa säätelevätkin hajureseptorit OBP-proteiinien toimiessa yleisinä kuljettajina (Löbel ym. 2002). Hyönteisillä on nisäkkäitä pienempi määrä hajureseptoreita (noin 70 *Drosophilalla*) ja suurempi määrä OBP-geenejä (noin 50 *Drosophilalla*) (Forêt ym. 2006). Tästä on päätelty, että hyönteisten OBP-proteiineilla voisi olla funktio myös hajumolekyylien valikoimisessa.

OBP- ja CSP-proteiinit suorittavat hajuaistimuksessa samankaltaisia tehtäviä (Pelosi ym. 2005), mutta eroavat toisistaan esimerkiksi laskostumiseltaan. OBP-proteiineille on tunnusomaista kuusi kysteiiniä, jotka muodostavat kolme disulfidisiltaa. CSP-proteiineilla taas on tavallisesti neljä kysteiiniä, jotka muodostavat kaksi disulfidisiltaa (Pelosi ym. 2005). OBP- ja CSP-geenien lukumäärä voi vaihdella suuresti hyönteislajien välillä. Tulimuurahaisella on 12 OBP-geeniä (Wurm ym. 2011), mehiläisellä 21 (Forêt ym. 2007) ja *Nasonia*-pistiäisellä 90 (Vieira ym. 2012). Muurahaisilla on 11-21 toimivaa CSP-geeniä per laji. Mehiläisellä niitä on kuusi ja *Drosophilalla* vain neljä (Kulmuni ym. 2013a).

OBP:t ja CSP:t muodostavat geeniperheitä, eli ajan kuluessa alkuperäisistä geeneistä on syntynyt useita kopioita genomiin. OBP- ja CSP-geeniperheisiin kuuluu hyvin konservoituneita geenejä, joiden ortologeja eli samankaltaisia, yhteisestä kantamuodosta peräisin olevia geenejä löytyy useilta sosiaalisilta hyönteisiltä. Hyönteislajien välisissä ortologisissa kemosensoirigeeneissä havaitaan puhdistavaa valintaa. Toisaalta perheisiin kuuluu evolutiivisesti nuorempia geenejä, jotka ovat spesifisiä tietyille linjoille tai lajeille ja joiden evoluutio on ollut dynaamisempaa. Jälkimmäisillä on korkea geenikuolleisuus ja uusien geenien syntyvyys ja geenien kopiolut eroavat lajien välillä. Yleisesti ottaen OBP:t muuntelevat enemmän lajien välillä ja lajin sisällä. CSP:t ovat konservoituneempia jopa kaukaisten lajien välillä (Pelosi ym. 2005).

OBP-sekvensseissä on muuntelua positiivisen valinnan seurauksena esimerkiksi kysteiinien määrässä. Menetetyt kysteiinit ovat yleensä juuri disulfidisidoksia muodostavia kysteiinejä (Vieira ym. 2012). Tietyt OBP:t ovat hyvin konservoituneita hyönteislajien välillä, mikä viittaa siihen, että niillä on tärkeä funktio. Osa OBP-geeneistä on linjaspesifisiä, kuten parasiittisten ampieisten ja hyttysten kaksoisdomeeniperheet (Vieira ym. 2012), jotka puuttuvat esimerkiksi muilta pistiäislinjoilta ja ovat siis syntyneet ampieisten ja hyttysten linjoissa.

Monet kemosensoirigeeniperheet ovat laajentuneet tietyissä hyönteislajeissa ja pienempi osa on konservoitunut lajien välillä (Kulmuni ym. 2013a). Muurahaisspesifisten, viimeaikoina duplikoituneiden, kemosensoirigeenien evoluutio on ollut nopeaa ja geenit ovat positiivisen valinnan alaisia (Kulmuni ym. 2013a). Näissä geeneissä luonnonvalinta on siis suosinut muuntelua

ja erilaistumista. Positiivinen valinta viittaa siihen, että näillä lajispesifisillä geeneillä voisi olla rooli lajiutumisen (Smadja ym. 2009) tai sopeutumisessa. Positiivisen valinnan alaisena olevat aminohapot geeneissä voivat mahdollisesti vaikuttaa siihen, mitkä signaalimolekyylit sitoutuvat kemosensoiriproteiineihin.

Hyönteisten OBP-geenit ilmentyvät haju- ja makuaistinelimissä (Pelosi ym. 2005). Osa OBP-geeneistä ilmentyy kuitenkin myös muissa kudoksissa tai vain tietyissä yksilönkehityksen vaiheissa (Forêt ym. 2006). Varsinkin lajispesifisten OBP- ja CSP-geenien on oletettu olevan mukana kemiallisessa kommunikaatiossa ja sopeutumisessa. Feromonit ja muut hajumolekyylit, joita OBP- ja CSP-proteiinit sitovat ja kuljettavat, voivat kehittyä nopeasti (Symonds ym. 2008). Jos OBP- ja CSP-geenit tunnistavat hajumolekyyliä spesifisesti, ne voisivat kehittyä koevoluutiossa hajumolekyylien kanssa. *Cerapachys biroi* -muurahaislajin ekspressioprofiilin mukaan suurin osa tuntosarvispesifisesti ekspressoituneista OBP- ja CSP-geeneistä kuuluu kuitenkin konservatiiviseen ryhmään (McKenzie ym. 2014). Lajispesifiset, nopeasti kehittyvät OBP- ja CSP-geenit sen sijaan ekspressoituvat useissa kudoksissa, eikä niiden tehtäviä tunneta. Tulos tukee päätelmää, että konservoituneilla OBP- ja CSP-geeneillä on tärkeä tehtävä hajuaistimuksessa, eivätkä ne sido hajumolekyyliä spesifisesti. Toisaalta OBP- ja CSP-geenien jako ekspression suhteen voi viitata siihen, että niillä on kaksoisrooli: ne saattavat toimia sekä signaalimolekyylien tunnistamisessa että niiden erityksessä (Pelosi ym. 2005).

Vaikka proteiinin yleisrakenne olisi konservoitunut, voi sille kehittyä lajispesifisiä funktioita, kuten hieman erilaisten ligandien sitominen eri lajeissa. Siksi tässä työssä tutkitaan, onko konservoituneissa geeneissä sekvenssimuuntelua muurahaislajien sisällä tai lajien välillä proteiinin rakennetta koodaavassa alueessa, joka voisi viitata jonkin asteiseen valikoivuuteen hajumolekyylien suhteen ja adaptiiviseen evoluutioon. Toisaalta tutkitaan muuntelua ei-koodaavalla alueella, esimerkiksi transkriptiofaktoreiden sitoutumiskohtien määrässä tai laadussa, sillä myös geeniekspression säätely voi olla evoluutiovoimien kohteena. Aiemmin ajateltiin, ettei valinta vaikuta kuin geenin koodaavaan osaan eli eksoneihin, koska ne määräävät proteiinin rakenteen. Viime aikoina on käynyt selväksi, että eksoneiden lisäksi valinta voi vaikuttaa myös ei-koodaaviin alueisiin (UTR-alueet ja intronit), esimerkiksi transkriptiofaktorien sitoutumiskohtiin. Transkriptiofaktorit säätelevät geenien ilmentymistä ja aktiivisuutta ja niiden sitoutumiskohtiin voi kohdistua esimerkiksi puhdistavaa valintaa, joka säilyttää sitoutumiskohdat samanlaisina. Geenisäätelyn muuntelulla oletetaan olleen tärkeä rooli esimerkiksi hyönteisten aitososiaalisuuden alkuperässä (Simola ym. 2013). Myös *Drosophila*-kärpäslajeilla on löydetty sekä

positiivisen että puhdistavan valinnan alaisia ei-koodaavia alueita (Haddrill ym. 2008, Andolfatto 2005).

2.3. Tutkimuksessa käytettävien hajugeenien funktiot

Tässä työssä käytettävät OBP- ja CSP-geenit ovat konservoituneita muurahaislajien välillä. Näiden geenien funktioiden oletetaan olevan samankaltaisia eri lajeissa. OBP-geenit on alun perin valittu mehiläisen ortologisten geenien ja muurahaisesta olemassa olevan sekvenssin perusteella. Tutkimuksessa käytettävistä lajeista ei ole genomisekvenssiä, joten alukkeet geenien monistamiseksi on suunniteltu sukulaislajien perusteella. Yksi tutkittavista geeneistä on OBP1 (tai ASP1), joka ilmentyy spesifisesti tuntosarvissa aikuisilla mehiläisillä (Forêt ym. 2006) ja sitoo kuningatarferomonia (Danty ym. 1999, Weng ym. 2014). Se ilmentyy vahvasti tuntosarvissa myös *Cerapachys biroi* -muurahaislajilla (McKenzie ym. 2014). Toinen tutkittava OBP-geeni on OBP10, joka ekspressoituu mehiläisellä kotelovaiheessa ja uusien aikuisten mehiläisten aivoissa (Forêt ym. 2006). *C. biroilla* se ekspressoituu useissa kudoksissa, mutta vahvemmin päässä kuin muualla (McKenzie ym. 2014).

Muurahaiset tunnistavat pesätoverinsa muista muurahaisista tai tunkeilijoista niiden erittämien kutikulaaristen hiilivetyjen (cuticular hydrocarbons, CHC) yhdistelmien perusteella (Singer 1998, Ozaki ym. 2005). Muurahaiset hyväksyvät pesätoverinsa ja torjuvat tai käyttäytyvät aggressiivisesti muita kohtaan. Tässä työssä tutkittava CSP7-geeni (AmelCSP1) ekspressoituu vahvasti tuntosarvissa usealla muurahaislajilla (Ozaki ym. 2005, McKenzie ym. 2014) ja sen on todettu sitovan pesätoverien tunnistuksessa käytettäviä kutikulaarisia hiilivetyjä *Camponotus japonicusella* (Ozaki ym. 2005). Aiemmin on esitetty, että juuri CSP7 voisi kehittyä positiivisen luonnonvalinnan alaisuudessa (Kulmuni ym. 2013a). Yhdyskunnat erittävät omia hajuyhdistelmiään ja tunkeilijat yrittävät kehittää samanlaisia yhdisteitä. Sopeutuakseen tilanteeseen ja erottaakseen pesätoverinsa tunkeilijoista, yhdyskunnan hajuyhdistelmä muuttuu edelleen ja hypoteesina on, että myös pesätoverin tunnistukseen liittyvä CSP7 kehittyisi valinnan alaisena. Toinen tutkittava CSP-geeni on CSP1 (AmelCSP4), joka ilmentyy mehiläisellä hajuaistimukseen liittyvissä kudoksissa. Ekspressiotaso on korkein tuntosarvissa, mutta se ekspressoituu myös päässä ja jaloissa (Forêt ym. 2007). Myös *C. biroilla* CSP1 ekspressoituu vahvimmin tuntosarvissa (McKenzie ym. 2014).

2.4. OBP-geenien yhteys muurahaisten sosiaaliseen muotoon

Tietyllä OBP-geenillä on havaittu yhteys muurahaisten sosiaaliseen muotoon. Selkeä yhteys tietyn geenin ja kompleksin käyttäytymissyndrooman, eli toisiinsa kytkeytyneiden käyttäytymispiirteiden, välillä oli ensin hämmästyttävää, sillä vain harvoissa tapauksissa yksi geeni saa aikaan niin suuren vaikutuksen useassa käyttäytymispiirteessä ja fenotyypissä.

Muurahaisten kaksi sosiaalista muotoa eli yhden ja usean kuningattaren yhdyskunnat eroavat toisistaan monessa suhteessa. Monogynisen eli yhden kuningattaren pesän ominaisuuksiin kuuluvat kuningattaren laajempi levittäytyminen (dispersaali), uuden pesän perustaminen ilman työläisten apua, lisääntyminen häälennolla (mating flight) sekä kuningattaren suurempi koko ja pidempi elinikä verrattuna polygyniseen pesään (Ross ym. 1995). OBP-geeneihin kuuluvassa Gp-9-geenissä on todettu löydetty eroja yhden kuningattaren ja monen kuningattaren yhdyskunnissa tulimuurahaisella (Krieger ym. 2002). Monogynisillä yhdyskunnilla on Gp-9-geenin B-alleeli ja polygynisillä on sekä B- että b-alleeleja. Tämän on ehdotettu vaikuttavan työläisten kykyyn säädellä kuningattaren määrää joko hyväksymällä tai tuhoamalla kuningatar sen erittämien kemiallisten signaalien perusteella. Geenin b-alleelissa on todettu positiivista valintaa, osittain sitoutumistaskun aminohapoissa (Krieger ym. 2005). Tästä on päätelty, että valinta on voinut ajaa muutoksia sitoutumistaskussa ja vaikuttaa Gp-9:n ligandin sitomiseen (Krieger ym. 2005).

Hiljattain huomattiin, että geeni sijaitsee alhaisen rekombinaation alueella, eikä Gp-9 näin ollen vaikuta monogynisten ja polygynisten yhdyskuntien eroihin yksin. Gp-9 on osa supergeeniä eli sosiaalikromosomia, johon kuuluu ainakin 616 geeniä (Wang ym. 2013). Tässä kromosomin alueessa on tapahtunut inversio, jolloin meioosi ei toimi normaalisti, eikä rekombinaatiota pääse tapahtumaan. Tässä supergeenissä ei siis ole rekombinaatiota monogynisiä ja polygynisiä pesiä erottavien B- ja b-alleelien välillä, mikä johtaa systemaattisiin eroihin monogynisten ja polygynisten yhdyskuntien välillä useassa muussakin geenissä. Gp-9:n rooli hajumolekyylien sitomisessa on myös epäselvä, koska Gp-9 ei ekspressoitu spesifisesti tuntosarvissa (Leal ym. 2008). Polygynisissä populaatioissa esiintyy sekä B:n suhteen homotsygoottisia että heterotsygoottisia yksilöitä. b-alleelin suhteen homotsygoottiset yksilöt sen sijaan ovat hyvin harvinaisia, koska b on yleensä resessiivinen letaali alleeli naarailla (Wang ym. 2013). *Formica selysin* genomista on löydetty vastaava sosiaalikromosomi eli kromosomialue, jossa ei tapahdu rekombinaatiota kahden alleelin välillä (Purcell ym. 2014). Tulimuurahasen ja *F. selysin* sosiaalikromosomit eivät kuitenkaan ole samaa evolutiivista alkuperää. *F. selysi* eroaa

tulimuurahaisesta myös siinä, että supergeenin polygynisiltä populaatioilta löytyvä alleeli Sp ei ole letaali ja voi esiintyä myös homotsygoottisena.

3. TYÖN TAVOITTEET

Tämän työn tavoitteena oli tutkia sekvenssimuuntelua ja mahdollista luonnonvalintaa kahdessa CSP- ja kahdessa OBP-geenissä (CSP1, CSP7, OBP1 ja OBP10). Nämä geenit ovat ortologisia usean muurahaislajin välillä, toisin sanoen sama geeni voidaan löytää usealta lajilta. Geenit on valittu alun perin niiden funktionaalisen kiinnostavuuden perusteella. Työssä tutkittavat kekomuurahaislajit ovat maailmanlaajuisesti uhanalaisia, mutta niitä on vaikea erottaa toisistaan perinteisin menetelmin. Jos kekomuurahaislajit eroavat toisistaan tutkituissa geeneissä, näitä geenejä olisi mahdollista käyttää lajintunnistuksessa ja suunniteltaessa suojelutoimenpiteitä.

Tutkimuskysymykset

1. Kuinka paljon sekvenssimuuntelua havaitaan läheisten lajien välillä, jotka ovat eronneet toisistaan viimeisen 500 000 vuoden aikana? Onko hajukommunikaation taustalla olevissa geeneissä lajin sisäistä muuntelua?
2. Mitkä evoluutiovoimat, luonnonvalinta vai satunnaisajautuminen, ovat muuntelun taustalla? Onko pesätoverin tunnistuksessa toimiva CSP7-geeni positiivisen valinnan alainen?
3. Jos luonnonvalintaa havaitaan, kohdistuuko se eksoneihin vai introneihin? Onko viitteitä siitä, että valinta kohdistuisi transkriptiofaktorien sitoutumiskohtiin?
4. Havaitaanko kahden eri sosiaalisen muodon (monogyninen ja polygyninen muoto) välillä systemaattisia eroja OBP- tai CSP-geeneissä, jotka voisivat viitata siihen, että nämä geenit vaikuttavat muurahaisyhdyskunnan sosiaaliseen rakenteeseen?

4. MATERIAALIT JA MENETELMÄT

4.1. Sekvenssiaineisto

Raaka sekvenssiaineisto saatiin valmiina ja se oli tuotettu Oulun yliopistossa professori Pekka Pamilon ryhmässä. Aineistossa on suorasekvensoitu neljän geenin (CSP1, CSP7, OBP1 ja OBP10) eksonit ja intronit. CSP1:n sekvensointiin on käytetty viittä aluketta, CSP7:n sekvensointiin kuutta aluketta, OBP1:n sekvensointiin 10 aluketta ja OBP10:n sekvensointiin seitsemää aluketta. Aineistoon kuuluu sekvenssejä seitsemästä *Formica*-muurahaislajista: *F. aquilonia* (tupsukekomuurahainen), *F. cinerea* (samettimuurahainen), *F. exsecta* (loviniskamuurahainen), *F. lugubris* (pohjankekomuurahainen), *F. polyctena* (kaljukekomuurahainen), *F. rufa* (punakekomuurahainen) ja *F. truncorum* (kantomuurahainen). Yhdestä geenistä on lisäksi *F. pratensis* (niitty-muurahainen) sekvenssejä. Yhteensä sekvenssiaineistoa on noin 270 yksilöstä. Alkuperäisten näytteiden lukumäärät ja keräyspaikat lajeittain sekä onnistuneiden näytteiden lukumäärät per geeni on lueteltu taulukossa 1. Aineisto on laaja ja mahdollistaa muuntelun ja valinnan tutkimisen eri taksonomisilla tasoilla (populaatio, laji, suku). Testeissä käytettiin ulkoryhmänä Hymenoptera Genome -tietokannan (Munoz-Torres ym. 2011) Ant Genome -portaalista (http://hymenopteragenome.org/ant_genomes/) haettua *Camponotus floridanus* -muurahaista (koko genomi: Cflo 3.3 Scaffold Assembly tai koodaava alue: Cflo OGSv3.3 cds) (Bonasio ym. 2010).

Taulukko 1. Alkuperäisten näytteiden lukumäärät ja keräyspaikat sekä onnistuneiden näytteiden lukumäärät konsensussekvenssien kasaamisen jälkeen geeneittäin.

Alkuperäiset näytteet			Onnistuneet näytteet (lkm)			
Laji	Yksilöiden lkm	Keräyspaikat	CSP1	CSP7	OBP1	OBP10
<i>F. aquilonia</i>	22	Venäjä, Irlanti, Skotlanti, Suomi, Skandinaavia	19	15	21	19
<i>F. cinerea</i>	97	Suomi (mono ja poly)	54	57	70	48
<i>F. exsecta</i>	64	Suomi (mono ja poly), Englanti, Romania, Venäjä, Ruotsi	14	34	53	49
<i>F. lugubris</i>	18	Irlanti, Skandinaavia, Venäjä	13	9	14	11
<i>F. polyctena</i>	26	Saksa, Skandinaavia, Venäjä	25	24	25	22
<i>F. pratensis</i>	4	Skandinaavia, Suomi, Venäjä	2	0	0	0
<i>F. rufa</i>	23	Saksa, Skandinaavia, Venäjä	21	16	21	5
<i>F. truncorum</i>	24	Suomi (mono ja poly)	11	24	24	24

4.2. Käytetyt menetelmät

Jokaisen lajin jokaiselle yksilölle koottiin oma konsensussekvenssi jokaisesta geenistä käyttäen ohjelmaa CodonCode Aligner (<http://www.codoncode.com>). Huonolaatuiset sekvenssit, joiden quality-arvo oli 0 tai hyvin alhainen, poistettiin. Mahdolliset heterotsygoottiset insertiot tai deleetiot etsittiin Find heterozygous indels -toiminnolla. Sekvensseistä poistettiin huonolaatuiset 5' ja 3' -päät Clips ends -toiminnolla. Yksilöille koottiin kontigit Assemble contigs -toiminnolla. OBP10-geenin kohdalla koottiin ensin yhden yksilön geeni liittämällä kaksi kontigia yhteen ja käytettiin tätä konsensussekvenssiä referenssinä muiden kontigien kokoamiseen. Osaan OBP10-geenin konsensussekvensseistä jäi noin 200 nt pituisia tai lyhyempiä aukkoja. Geeneissä on pitkiä A- ja T-toistojaksoja, joiden jälkeinen sekvenssi oli vaikeasti luettavaa. Tällaiset heterotsygoottisilta insertioilta tai deleetioilta vaikuttavat jaksot tulkittiin johtuvaksi PCR-vaiheen virheellisestä monistumisesta, eikä niitä merkitty konsensussekvensseihin. Näiden alueiden sekvenssi luettiin toisesta, parempilaatuisemmasta juosteesta. Noin puolesta sekvensoiduista yksilöistä saatiin tuotettua konsensussekvenssit. Loppujen yksilöiden sekvensseistä joko kaikki tai osa oli niin huonolaatuisia, etteivät ne yhdistyneet kontigeiksi.

4.2.1. Geenien rinnastus ja annotointi

Tässä työssä kootut sekvenssit rinnastettiin geeneittäin usean sekvenssin rinnastuksella MAFFT-ohjelmalla (<http://mafft.cbrc.jp/alignment/software>) (Katoh ym. 2013) käyttäen oletusparametreja. Rinnastukset tarkastettiin ja niitä muokattiin Jalview:lla (<http://www.jalview.org>) (Waterhouse ym. 2009) ja Bioedit:illä (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>) (Hall 1999). Geenien eksoni-introni-rajat määriteltiin lähimmän sekvensoidun muurahaisen *Camponotus floridanuksen* ja mehiläisen *Apis mellifera*n cDNA:iden ja aminohapposekvenssien perusteella käyttäen apuna Genewise (<http://www.ebi.ac.uk/Tools/psa/genewise>) (Birney ym. 2004) ja Expasy Translate (<http://web.expasy.org/translate>) (Artimo ym. 2012) -työkaluja. Genewise:n avulla nukleotidisekvenssi voidaan rinnastaa aminohapposekvenssin kanssa, jolloin myös todennäköiset introni-eksonirajat voidaan tunnistaa. Expasy Translate kääntää nukleotidisekvenssin aminohapposekvenssiksi kaikissa mahdollisissa lukuraameissa. Heterotsygoottisia SNP:ejä sisältävät sekvenssit jaettiin alleeleiksi PHASE-ohjelmistolla (oletusparametrit) (Stephens ym. 2001; Stephens ym. 2005). PHASE:n syötteet tehtiin fasta-tiedostoista seqPhase-työkalulla

(<http://seqphase.mpg.de/seqphase>) (Flot ym. 2010). Heterotsygoottisten SNP:ien erottelun lisäksi PHASE arvioi puuttuvaksi merkittyjen alueiden todennäköisimmät alleelit.

4.2.2. Sekvenssimuuntelua ja lajien välisiä eroja kuvaavat analyysit

Tässä työssä tutkittiin läheisiä muurahaislajeja ja ortologisia geenejä, jotka ovat konservoituneita muurahaislajien välillä. Muuntelun määrää ja lajien välisiä suhteita tutkittiin eri menetelmin, joista yksi oli fylogeneettinen puu. Fylogeneettinen puu tehtiin MEGA:lla (<http://www.megasoftware.net>) (Tamura ym. 2013) neighbor joining -menetelmällä fasta-muotoisesta sekvenssirinnastuksesta, joka sisälsi koodaavan ja ei-koodaavan alueen kaikista neljästä geenistä. Oksanhaarojen tilastolliset uskottavuudet testattiin bootstrap-analyysillä, 1000 uudelleenotannalla. Bootstrap-analyysi kertoo, kuinka hyvin näyteotos tukee piirrettyä puuta. Tämä tehdään ottamalla aineistosta uudelleenotoksia ja tilastoimalla kuinka usein alkuperäisen puun oksanhaara esiintyy näiden uudelleenotantojen tuottamissa puissa. Tuloksena saatua newick-muotoista puuta muokattiin FigTree:ssä (<http://tree.bio.ed.ac.uk/software/figtree>) niin, että puussa näkyy vain toisistaan erottuvat lajit yksittäisten yksilöiden sijaan. Puuttuvien alueiden ja deleetioiden/insertioiden käsittelyyn valittiin parittainen deleetio.

Lajien sisäistä muuntelua verrattuna lajien väliseen muunteluun visualisoitiin pääkoordinaattianalyysillä (Principal coordinate analysis, PCoA). Siinä valittiin kaikki yksilöt, joista on sekvenssi ainakin yhdestä geenistä (223 yksilöä) ja liitettiin kaikkien neljän geenin sekvenssit yhteen. Kaikkien yksilöiden väliset parittaiset geneettiset etäisyydet laskettiin MEGA:lla maximum likelihood-menetelmällä. Puuttuvien alueiden ja deleetioiden/insertioiden käsittelyyn valittiin parittainen deleetio. PCoA tehtiin parittaisten etäisyyksien perusteella GenALEX-ohjelmistolla (Peakall ym. 2006, 2012). Jos lajit erottuvat toisistaan sekvenssiaineiston perusteella, ne näkyvät myös PCoA-kuvassa selkeästi omina ryhminään.

F_{ST} -arvo (fixation index) mittaa geneettisestä rakenteesta johtuvaa populaatioiden erilaisuutta. F_{ST} -arvo on osapopulaation odotettu heterotsygotia verrattuna koko populaation heterotsygotiaan. Testisuureen arvo vaihtelee välillä 0 - 1. F_{ST} :n arvo 0 tarkoittaa, että populaatiot risteytyvät vapaasti ja alleelifrekvensseihin vaikuttaa vain satunnainen geneettinen ajautuminen. Kun osapopulaatioiden väliset alleelifrekvenssit muuttuvat, koko populaatiossa on alimäärä heterotsygootteja verrattuna osapopulaatioihin ja F_{ST} kasvaa. Korkea ja merkitsevä F_{ST} -arvo tarkoittaa, että osapopulaatiot eivät risteydy vapaasti keskenään. F_{ST} :n korkein teoreettinen arvo 1

saavutetaan, kun kaikki populaatiot ovat fiksoituneet jonkin alleelin suhteen. F_{ST} -arvo on kuitenkin yleensä huomattavasti alle 1. Esimerkiksi F_{ST} -arvo 0,10 tarkoittaa, että 90 % muuntelusta on populaation sisällä ja vain 10 % populaatioiden välillä. Wrightin mukaan 0 – 0,05 suuruisen F_{ST} -arvon voi tulkita merkitsevän vähäistä eroavuutta populaatioiden välillä, 0,05 – 0,15 kohtuullista eroavuutta, 0,15 – 0,25 suurta eroavuutta ja yli 0,25 hyvin suurta eroavuutta (Hartl ym. 2007: s. 283).

Lajin sisällä populaatioiden väliset parittaiset F_{ST} -arvot ja niiden merkitsevyys mitattiin Arlequin 3.5 -ohjelmalla (Excoffier ym. 2010) käyttäen etäisyysmenetelmää ja 1000 uudelleenotantaa. Mittaukset tehtiin *F. cinerean*, *F. exsectan* ja *F. truncorum*in monogynisille ja polygynisille populaatioille, koska näistä ryhmistä ja lajeista on suurimmat sekvenssiaineistot. Tällä menetelmällä selvitettiin voidaanko populaatioita yhdistää suuremmiksi populaatioiksi Tajiman D ja Fu ja Li -evoluutiotestejä varten. Nämä testit vaativat syötteeksi mahdollisimman suuren otoksen panmiktisesta populaatiosta. Parittaiset F_{ST} -arvot laskettiin käyttäen CSP1-, CSP7-, OBP1- ja OBP10-geenien sekvenssejä yhteen liitettyinä.

Nukleotididiversiteetti (π , π) on keskimääräinen määrä nukleotidieroja per nukleotidikohta kahden sekvenssin välillä. Nukleotididiversiteetti laskettiin DnaSP-ohjelmistolla (Librado ym. 2009) *F. cinerean*, *F. exsectan*, *F. truncorum*in ja muun *F. rufa* -ryhmän jokaiselle geenille sisältäen sekä koodaavan että ei-koodaavan osan.

Fiksoituneet erot (fixed differences) lajien välillä laskettiin DnaSP-ohjelmistolla. Fiksoitunut ero kahden lajin välisessä vertailussa on kohta, jossa kaikki yhden lajin sekvenssien nukleotidit eroavat toisen lajin nukleotideista. Analyysi tehtiin erikseen kaikille geeneille. Tuloksissa eroteltiin kaikki fiksoituneet erot ja vain eksoneihin sijoittuvat fiksoituneet erot. Parittaiset vertailut tehtiin kaikkien kahdeksan lajin välillä. Tuloksista näkee lajien erilaistumisen lisäksi sen, antavatko eri geenit lajien suhteista samaa tietoa.

F. cinerean, *F. exsectan* ja *F. truncorum*in monogynisistä ja polygynisistä sekvenssiaineistoista etsittiin SNP:ejä, jotka esiintyvät vain polygynisissä populaatioissa. Tuloksiin listattiin geenit vain niistä lajeista, joilla vähintään kaksi erillistä kyseisen kaltaista SNP:iä vähintään kahdella yksilöllä. *F. cinerean* tuloksissa merkittiin lisäksi, esiintyvätkö SNP:it molemmissa vai pelkästään toisessa polygynisessä populaatiossa.

4.2.3. Evoluutioanalyysit

Evoluutiovoimien analysointi tehtiin neljällä eri testillä: McDonald-Kreitman (MK), Tajiman D, Fu ja Li sekä MFDM. Kaikki testit olettavat, ettei tutkittavassa alueessa ole tapahtunut rekombinaatiota. Muuten testit perustuvat osittain eri olettamuksiin ja ovat herkkiä erityyppiselle valinnalle. Tajiman D (Tajima 1989) ja Fu ja Li (Fu ym. 1993) -testeillä mitataan, onko aineistossa viimeaikaista tasapainottavaa tai positiivista/negatiivista valintaa. Näillä testeillä ei siis voi erottaa positiivista ja negatiivista valintaa toisistaan. MK-testi (McDonald ym. 1991) tunnistaa aiemman lajien välillä tapahtuneen positiivisen tai negatiivisen valinnan, mutta ei tasapainottavaa valintaa. MFDM-testi (Li 2011) tunnistaa vain viimeaikaisen positiivisen valinnan. Fu ja Li -testi hylkää neutraalin nollahypoteesin Tajiman D -testiä herkemmin, osittain sen takia, että Fu ja Li -testissä käytetään ulkoryhmää. Fu ja Li -testi on tehokas taustavalinnan eli puhdistavan valinnan havaitsemisessa, Tajiman D taas havaitsee paremmin geneettisen hitchhikingin eli positiivisen valinnan tai populaatioekspansion (Fu 1997). Myös MK ja MFDM-testeihin tarvitaan ulkoryhmä. MK ja MFDM-testit eivät ole herkkiä populaatorakenteelle toisin kuin Tajiman D ja Fu ja Li -testit (Eilertson ym. 2012, Li 2011). Kaksi ensimmäistä testiä voidaan siis tehdä koko lajille, mutta kaksi jälkimmäistä testiä tehdään lajien populaatioille.

McDonald-Kreitman -testi (McDonald ym. 1991) perustuu koodaavan alueen synonyymisten ja ei-synonyymisten mutaatioiden suhteeseen lajin sisällä ja lajien välillä. Tutkittavasta lajista tarvitaan useiden yksilöiden sekvenssit, ulkoryhmälajista riittää yksi sekvenssi. Neutraali malli, jossa geeni ei ole valinnan alainen, olettaa että synonyymisten ja ei-synonyymisten mutaatioiden suhde on sama lajin sisällä ja lajien välillä. Näiden merkitsevä eroaminen toisistaan voi olla merkki positiivisesta tai puhdistavasta valinnasta. MK-testi tehtiin DnaSP-ohjelmistolla (Librado ym. 2009) jokaiselle kahdeksalle *Formica*-lajille käyttäen *Camponotus floridanus* ulkoryhmänä. DnaSP ei ota analyyseissä huomioon kohtia, joissa jollakin aineiston yksilöllä on puuttuva nukleotidi tai deleetio. Syötteenä käytettiin aineistoa, jossa sekvenssien puuttuvia alueita oli muokattu vastaamaan kyseisen lajin yleisintä alleelimuotoa. Tällä tavalla saatiin mahdollisimman suuri osa aineiston muuntelevista kohdista käyttöön. MK-testi ei perustu alleelifrekvensseihin, vaan varioivien kohtien lukumääriin, joten tämä aineiston muokkaus ei vaikuta MK-testin tuloksiin. Testin tuloksena saadaan MK-taulukko (Taulukko 2), tuloksen merkitsevyys, neutraalisuusindeksi (neutrality index, NI) ja α . Jos lajien välisten ei-synonyymisten mutaatioiden suhde synonyymisiin mutaatioihin on merkitsevästi suurempi kuin vastaava lajin sisäinen suhde, tulos viittaa positiiviseen valintaan. Päinvastainen tulos, jossa lajin sisäinen suhde on suurempi kuin lajien välinen suhde, viittaa puhdistavaan valintaan. Jos suhteet ovat yhtä suuret, neutraalia nollahypoteesia ei voida kumota.

Sama nähdään neutraalisuusindeksistä (Rand ym. 1996) (Kaava 1). $NI > 1$ (tai $\alpha < 0$) tarkoittaa ylimääräistä aminohappomuuntelua lajin sisällä verrattuna lajien väliseen variaatioon eli mahdollista puhdistavaa valintaa. $NI < 1$ (tai α välillä 0 - 1) tarkoittaa ylimääräistä aminohappovariaatiota lajien välillä eli mahdollista positiivista valintaa. Tulos $NI = 1$ (tai $\alpha=0$) taas kertoo, ettei nollahypoteesia voida kumota. Pullonkaula eli hetkellinen pieni populaatiokoko, lievästi haitalliset mutaatiot, epävaka demografia kuten vaihteleva populaatiokoko tai sukupuolijakauma, pienet näytekoot tai hiljattain eronneet lajit voivat vääristää tulosta. MK-testi sietää kuitenkin demografiaa hyvin verrattuna moniin muihin valinnan testeihin.

	Lajien välinen variaatio	Lajin sisäinen polymorfismi
Ei-synonyyminen	D_N	P_N
Synonyyminen	D_S	P_S

Taulukko 2. MK-testi vertaa lajien välisten ja lajin sisäisten ei-synonyymisten ja synonyymisten mutaatioiden suhdetta.

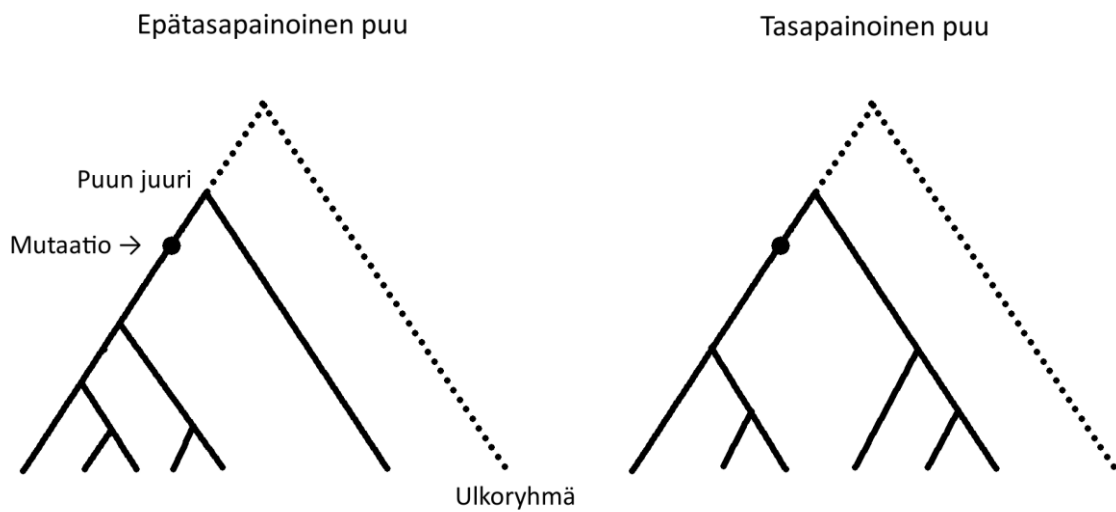
$$NI = \frac{P_N \times D_S}{D_N \times P_S} = \frac{P_N}{D_N} / \frac{P_S}{D_S}$$

Kaava 1. NI:n kaava MK-tilaukosta.

Tajiman D (Tajima 1989) ja Fu ja Li (Fu ym. 1993) -testejä käytetään valinnan tunnistamiseen populaatiosekvenssiaineistosta. Testien nollahypoteesina on, että mutaatiot ovat neutraaleja. Nollasta poikkeavat merkitsevät tulokset viittaavat joko valintaan tai demografisiin tapahtumiin kuten populaation koon muutoksiin, koska molemmat jättävät samankaltaisia signaaleja. Tajiman D -statistiikka testaa, sopiiko populaation alleelifrekvenssispektri neutraaliin malliin, vertaamalla segregoituvien kohtien lukumäärää keskimääräiseen kahden sekvenssin nukleotidierojen määrään. Nämä korreloivat keskenään, paitsi jos neutraalin mallin oletukset eivät ole voimassa eli jos geeni on esimerkiksi valinnan alainen. Erot näiden kahden välillä johtuvat siitä, että polymorfisten nukleotidien frekvenssit vaikuttavat sekvenssien keskimääräisten parittaisten erojen määrään, mutta eivät segregoituvien kohtien lukumäärään. Parittaisten erojen ylimäärä johtaa positiiviseen Tajiman D -tulokseen ja viittaa muuntelun ylläpitämiseen ja tasapainottavaan valintaan tai populaation pienentymiseen. Parittaisten erojen ja muuntelun vähäisyys tuottaa negatiivisen Tajiman D -tuloksen ja voi viitata joko positiiviseen tai negatiiviseen valintaan tai kasvavaan populaatioon. Ei-merkitsevä Tajiman D tulos voi viitata neutraalin mallin sijasta myös kaukaisempaan valinnan pyyhkäisyyn tai heikkoon valintaan.

Fu ja Li -testi (Fu ym. 1993) taas vertaa viimeaikaisten ja aikaisempien mutaatioiden määrää tutkimalla mutaatioita geenipuun ulkohaaroissa verrattuna puun sisäosiin. Toisin kuin Tajiman D, Fu ja Li -testi käyttää alkoryhmää. Merkitsevä negatiivinen Fu ja Li -testin tulos saadaan, jos puun ulko-oksissa on odotettua enemmän mutaatioita. Tämä indikoi mahdollista puhdistavaa valintaa, koska resessiiviset haitalliset alleelit segregoituvat alhaisella frekvenssillä eikä niistä päästä eroon, tai positiivista valintaa, koska mutaatioiden oletetaan silloin olevan nuoria. Merkitsevä positiivinen tulos taas saadaan, kun puun sisäosissa on odotettua enemmän mutaatioita, jolloin muuntelua siis ylläpidetään, ja joka viittaa tasapainottavaan valintaan. Ei-merkitsevä tulos tarkoittaa, että neutraalia nollahypoteesia ei voida kumota. Testit tehtiin DnaSP-ohjelmistolla. Fu ja Li -testissä käytettiin alkoryhmänä *Camponotus floridanus*. Aineistona käytettiin kolmea *Formica*-lajia, joista oli suurimmat populaatioaineistot saatavilla: *F. cinerea*, *F. exsecta* ja *F. truncorum*. *F. cinerean* testit tehtiin erikseen kahdelle monogyniselle ja kahdelle polygyniselle populaatiolle. *F. exsectan* aineistoon yhdistettiin monogyniset ja polygyniset populaatiot yhdeksi populaatioksi. *F. truncorumin* testit tehtiin erikseen monogyniselle ja polygyniselle populaatiolle. Aineistossa kaikki alun perin puuttuvat nukleotidit jätettiin puuttuviksi. DnaSP ei kuitenkaan ota evoluutioanalyysiin mukaan kohtia, joissa jollakin sekvenssillä on puuttuva tai deletoitunut nukleotidi. Tämä vähentää analysoitavia kohtia. Sen vuoksi osa sekvensseistä, joissa oli pitkiä puuttuvia alueita, poistettiin ennen analyysiä. Testeihin käytettiin sekä koodaavaa että ei-koodaavaa osaa geeneistä. Testit tehtiin sekä koko geeneille että 100 nt osille sliding window -menetelmällä.

MFDM (Maximum frequency of derived mutations) -testi (Li 2011) testaa viimeaikaista positiivista valintaa. Toisin kuin monet muut valinnan testit, MFDM-testi ei perustu alleelifrekvenssispektriin, vaan fylogeniaan. MFDM-testi tunnistaa uusien mutaatioiden maksimifrekvenssin epätasapainoisesta puutopologiasta, joka viittaa positiiviseen valintaan. Epätasapainoisessa puussa yhteen sisäoksanhaaraan liittyy monta ulko-oksaa ja toiseen sisäoksaan ei liity ollenkaan ulko-oksa (Kuva 2). Tällainen puu syntyy, jos positiivinen valinta on nostanut yhden alleelin frekvenssin korkeaksi ja kyseiseen alleleliin on kertynyt uusia mutaatioita.



Kuva 2. Epätasapainoinen ja tasapainoinen puu. Epätasapainoisessa puussa vasempaan sisäoksaan liittyy useita ulko-oksia, mutta oikeaan sisäoksaan ei yhtäkään. Tasapainoisessa puussa molempiin sisäoksiin liittyy ulko-oksia. (Kuvan malli: Li ym. 2012)

Populaatiorakenne tai populaation koon muutokset kuten pullonkaula tai populaatioekspansio eivät vaikuta MFDM:n tulokseen toisin kuin moniin muihin valinnan testeihin (Li ym. 2012). Demografia vaikuttaa koko genomiin ja valinnan taas oletetaan vaikuttavan vain paikallisesti. Usein populaatiodemografian poissulkemiseksi on analysoitava koko genomi tai useita alueita, jolloin nähdään erottuuko tietty alue muista. MFDM-testiin riittää kuitenkin yksi lokus. MFDM-testiin tarvitaan vähintään 21 diploidia yksilöä merkitsevyystasoon $p = 0,05$. MFDM -testi tehtiin *F. cinerean*, *F. exsectan* ja *F. truncorum* kaikille yksilöille. Migrant detector -parametreina annettiin yksi yksilö kaikista pienimmistä populaatioista. Ulkoryhmänä käytettiin toista *Formica*-lajia: *F. cinerean* ulkoryhmänä oli *F. exsecta* ja *F. exsectan* ja *F. truncorum* ulkoryhmänä *F. cinerea*.

Geenin sisäiset rekombinaatiotapahtumat voivat vaikuttaa evoluutioanalyysien tuloksiin, testistä riippuen joko lisäämällä väärän positiivisen tuloksen mahdollisuutta tai tekemällä testistä konservatiivisen. Esimerkiksi MFDM-testi antaa herkemmin väärän positiivisen tuloksen, kun geenin sisällä on mahdollista rekombinaatiota. Jos MFDM havaitsee sekvenssissä rekombinaatiotapahtumia, merkitsevyysrajaa tiukennetaan Bonferroni-korjauksella. Korjaus tekee testistä konservatiivisemman, jolloin väärä negatiivinen tulos on todennäköisempi. MFDM:n käyttämän rekombinaatiotestauksen lisäksi rekombinaatiota testattiin HyPhy-ohjelmiston (Pond ym. 2005a) SBP (Single Break Point) -rekombinaatiotestillä (Kosakovsky Pond ym. 2006)

Datamonkey-serverillä (Delpont ym. 2010, Pond ym. 2005b) (<http://www.datamonkey.org>) käyttäen oletusparametreja. SBP-testin tulokset kertovat onko sekvenssirinnastuksessa eri kriteerien perusteella todisteita rekombinaatiosta ainakin yhdessä tutkituista mahdollisista rekombinaatiokohdista. SBP:n kriteerit ovat AIC (Akaike Information Criterion), AIC pienelle aineistolle (AICc) ja BIC, joista pienelle näytemäärälle suositellaan AIC/AICc -kriteerejä. BIC on näistä konservatiivisin. Tulos kertoo, kuinka paljon oletusmalli ja aineisto eroavat toisistaan, eli onko aineistossa todennäköisiä rekombinaatiokohtia. Testi tehtiin jokaiselle neljälle geenille ja erikseen *F. cinerea*, *F. exsecta* ja *F. truncorum* -lajeille. Aineistossa oli mukana sekä koodaava että ei-koodaava alue.

Geenien ei-koodaavista osista ennustettiin transkriptiofaktorien sitoutumiskohtia. Hyvin harvoista sekvensseissä oli mukana promoottorialueita, joten analyysiin voitiin käyttää vain introneita. Transkriptiofaktorit saattavat sitoutua geenejä edeltävien promoottorialueiden lisäksi myös introneihin geenin sisällä. Tajiman D ja Fu ja Li -testien tuloksina saatiin geenien alueita, jotka ovat mahdollisesti valinnan alaisia. Tällaisista introneista ennustettiin mahdollisia transkriptiofaktoreiden sitoutumiskohtia ALGGEN-serverin (<http://alggen.lsi.upc.es>) PROMO-ohjelmalla (Farré ym. 2003). PROMO käyttää TRANSFAC-tietokannan transkriptiofaktoreita. Syötteeseen valittiin 1-4 sekvenssiä per laji jokaisesta geenistä. Tulosten rajaamiseksi valittiin vain hyönteisten transkriptiofaktorit ja vaadittiin sitoutumiskohtien sekvenssien täyttä identtisyyttä. Sitoutumiskohtien sekvensseissä voi kuitenkin olla jonkin verran muuntelua, joten mahdollisia sitoutumiskohtia on luultavasti enemmän. Kaikki sitoutumiskohdan sekvenssin nukleotidit eivät ole välttämättömiä transkriptiofaktorin sitoutumiselle (Weingarten-Gabbay ym. 2014). Toisaalta laskennallisten transkriptiofaktoreiden sitoutumiskohtien ennustusanalyysillä on huono selektiivisyys. Ennustettuihin sekvensseihin siis sitoutuu transkriptiofaktoreita todellisuudessa harvoin (Wasserman ym. 2004). Vaikka sekvenssin spesifisyys olisi hyvin korkea, transkriptiofaktori ei välttämättä sitoudu siihen (Weingarten-Gabbay ym. 2014). Tällaisia analyysijä voi kuitenkin käyttää transkriptiofaktoreiden sitoutumiskohtien tutkimisen ensimmäisenä vaiheena. Varsinaisten johtopäätösten tekemiseen tarvitaan aina kokeellisia tutkimuksia.

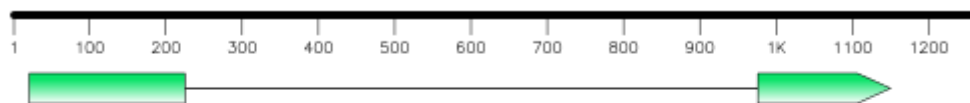
5. TULOKSET

5.1. Geenien rakenne ja lokus

CSP-geeneissä on kaksi eksonia ja OBP-geeneissä viisi eksonia. Geenien pituudet ovat 900 nt (CSP1), 1100 nt (CSP7), 1300 nt (OBP1) ja 1800 nt (OBP10). Geenien rakenteet näkyvät kuvissa 3a - d ja tarkat eksoni-introni-rajakohdat taulukossa 3. CSP7-geenin promoottorialue ja viimeisen eksonin loppuosa puuttuvat melkein kaikista näytesekvensseistä. Monista sekvensseistä puuttuu OBP10-geenin keskiosa.



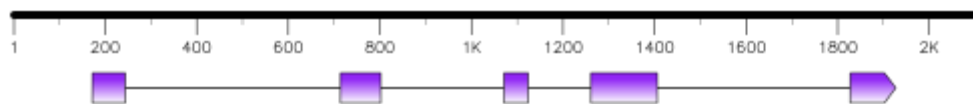
Kuva 3a. CSP1-geenin rakenne. Eksonit on merkitty sinisellä, juosteen suunta nuolella ja introni mustalla viivalla. Geenin pituus ensimmäisen eksonin alusta toisen eksonin loppuun, introni mukaan lukien, on noin 900 nt.



Kuva 3b. CSP7-geenin rakenne. Eksonit on merkitty vihreällä, juosteen suunta nuolella ja introni mustalla viivalla. Geenin pituus ensimmäisen eksonin alusta toisen eksonin loppuun, introni mukaan lukien, on noin 1100 nt.



Kuva 3c. OBP1-geenin rakenne. Eksonit on merkitty punaisella, juosteen suunta nuolella ja intronit mustalla viivalla. Geenin pituus ensimmäisen eksonin alusta viimeisen eksonin loppuun, intronit mukaan lukien, on noin 1250 nt.



Kuva 3d. OBP10-geenin rakenne. Eksonit on merkitty violetilla, juosteen suunta nuolella ja intronit mustalla viivalla. Geenin pituus ensimmäisen eksonin alusta viimeisen eksonin loppuun, intronit mukaan lukien, on noin 1800 nt.

Taulukko 3. CSP- ja OBP-geenien eksoneiden alku- ja loppukohdat. E1, E2 ja niin edelleen viittaavat eksoneihin.

Geeni	E1 alku	E1 loppu	E2 alku	E2 loppu	E3 alku	E3 loppu	E4 alku	E4 loppu	E5 alku	E5 loppu
CSP1	39	237	726	961	-	-	-	-	-	-
CSP7	21	224	976	1149	-	-	-	-	-	-
OBP1	5	65	292	376	776	864	942	1044	1147	1268
OBP10	172	243	714	801	1071	1123	1261	1405	1827	1926

Kaikki tutkittavat geenit sijaitsevat *Camponotus floridanuksen* genomisekvenssin eri scaffoldeissa, eivätkä näin ollen sijaitse kovin lähellä toisiaan (Taulukko 4), eivätkä siis luultavasti ole kytkeytyneitä. Jos eri geenien siis havaitaan olevan samankaltaisten evoluutiovoimien alaisena, syynä ei ole se, että geenit olisivat kytkeytyneitä toisiinsa ja valinta yhdessä geenissä voisi saada aikaan saman valinnan signaalin toisessakin geenissä.

Geeni	Scaffold	Nukleotidi- koordinaatit
CSP1	942	487076-488003
CSP7	1433	258591-259623
OBP1	28	211108-212148
OBP10	2425	86399-88074

Taulukko 4. CSP1-, CSP7-, OBP1- ja OBP10-geenien lokukset *Camponotus floridanuksen* genomissa.

5.2. Sekvenssimuuntelu lajeissa

5.2.1. Fiksoituneet muutokset lajien välillä

Kaikkien lajien väliset koodaavan ja ei-koodaavan alueen parittaiset fiksoituneet muutokset näyttävät, että *F. cinerea* ja *F. exsecta* eroavat muista lajeista jokaisen geenin suhteen (Taulukot 5a - d). Lisäksi *F. truncorum* eroaa yhdellä fiksoituneella erolla kaikista muista *F. rufa* -ryhmän lajeista OBP1-geenin suhteen (Taulukko 5c), *F. rufasta* CSP1-geenin suhteen (Taulukko 5a) ja *F. rufasta* ja *F. aquiloniasta* CSP7-geenin suhteen (Taulukko 5b). *F. rufa* eroaa *F. aquiloniasta* ja *F. lugubriksesta* CSP7-geenin suhteen (Taulukko 5b). *F. pratensiksesta* on sekvenssejä vain CSP1-geenistä ja *F. pratensis* eroaa siinä *F. aquiloniasta* ja *F. rufasta* (Taulukko 5a). Huomattavaa on, että *F. aquilonian*, *F. lugubriksen* ja *F. polychtenan* välillä ei ole fiksoituneita eroja yhdessäkään geenissä. Suurin osa fiksoituneista eroista on ei-koodaavalla alueella. *F. rufa* -ryhmän lajien välillä ei ole fiksoituneita eroja eksoneissa muissa geneeissä kuin OBP1:ssä, jossa *F. truncorumilla* on yksi fiksoitunut ero muiden *F. rufa* -ryhmän lajien kanssa. DnaSP ei ota huomioon deleetioita ja insertioita. *F. exsectan* ja muiden lajien välillä on indeleitä geenissä OBP10 ja *F. exsectan* sekä *F. cinerean* ja muiden lajien välillä on indeleitä geenissä CSP7.

CSP1

	<i>Aq.</i>	<i>Cin.</i>	<i>Ex.</i>	<i>Lug.</i>	<i>Pol.</i>	<i>Pra.</i>	<i>Ruf.</i>
<i>F. aquilonia</i>							
<i>F. cinerea</i>	5 (2)						
<i>F. exsecta</i>	1 (0)	6 (2)					
<i>F. lugubris</i>	0	5 (2)	1 (0)				
<i>F. polychtena</i>	0	5 (2)	1 (0)	0			
<i>F. pratensis</i>	1 (0)	6 (2)	1 (0)	0	0		
<i>F. rufa</i>	0	6 (2)	4 (0)	0	0	1 (0)	
<i>F. truncorum</i>	0	7 (2)	3 (0)	0	0	0	1 (0)

Taulukko 5a. Parittaiset fiksoituneet muutokset kaikkien lajien välillä geenissä CSP1. Ensimmäinen luku kertoo kaikki muutokset ja sulussa oleva luku eksoneiden muutokset.

CSP7

	<i>Aq.</i>	<i>Cin.</i>	<i>Ex.</i>	<i>Lug.</i>	<i>Pol.</i>	<i>Ruf.</i>
<i>F. aquilonia</i>						
<i>F. cinerea</i>	7 (2)					
<i>F. exsecta</i>	13 (3)	10 (3)				
<i>F. lugubris</i>	0	7 (2)	13 (3)			
<i>F. polychtena</i>	0	7 (2)	13 (3)	0		
<i>F. rufa</i>	2 (0)	9 (2)	14 (3)	1 (0)	0	
<i>F. truncorum</i>	1 (0)	8 (2)	13 (3)	0	0	1 (0)

Taulukko 5b. Parittaiset fiksoituneet muutokset kaikkien lajien välillä geenissä CSP7. Ensimmäinen luku kertoo kaikki muutokset ja sulussa oleva luku eksoneiden muutokset.

OBP1

	<i>Aq.</i>	<i>Cin.</i>	<i>Ex.</i>	<i>Lug.</i>	<i>Pol.</i>	<i>Ruf.</i>
<i>F. aquilonia</i>						
<i>F. cinerea</i>	4 (1)					
<i>F. exsecta</i>	6 (2)	5 (2)				
<i>F. lugubris</i>	0	3 (1)	6 (2)			
<i>F. polyclena</i>	0	3 (1)	5 (2)	0		
<i>F. rufa</i>	0	2 (1)	4 (2)	0	0	
<i>F. truncorum</i>	1 (1)	4 (2)	6 (3)	1 (1)	1 (1)	1 (1)

Taulukko 5c. Parittaiset

fiksoituneet muutokset kaikkien lajien välillä geenissä OBP1.

Ensimmäinen luku kertoo kaikki muutokset ja suluissa oleva luku eksoneiden muutokset.

OBP10

	<i>Aq.</i>	<i>Cin.</i>	<i>Ex.</i>	<i>Lug.</i>	<i>Pol.</i>	<i>Ruf.</i>
<i>F. aquilonia</i>						
<i>F. cinerea</i>	10 (2)					
<i>F. exsecta</i>	8 (1)	10 (1)				
<i>F. lugubris</i>	0	10 (2)	8 (1)			
<i>F. polyclena</i>	0	9 (2)	7 (1)	0		
<i>F. rufa</i>	0	10 (2)	8 (1)	0	0	
<i>F. truncorum</i>	0	8 (2)	6 (1)	0	0	0

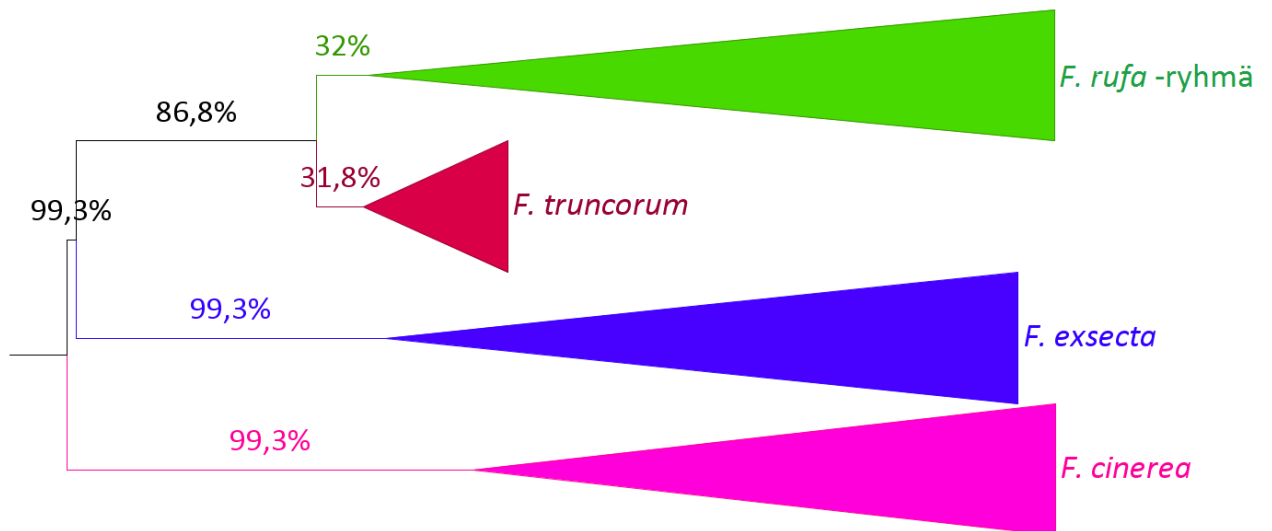
Taulukko 5d. Parittaiset

fiksoituneet muutokset kaikkien lajien välillä geenissä OBP10.

Ensimmäinen luku kertoo kaikki muutokset ja suluissa oleva luku eksoneiden muutokset.

5.2.2. Fylogenia

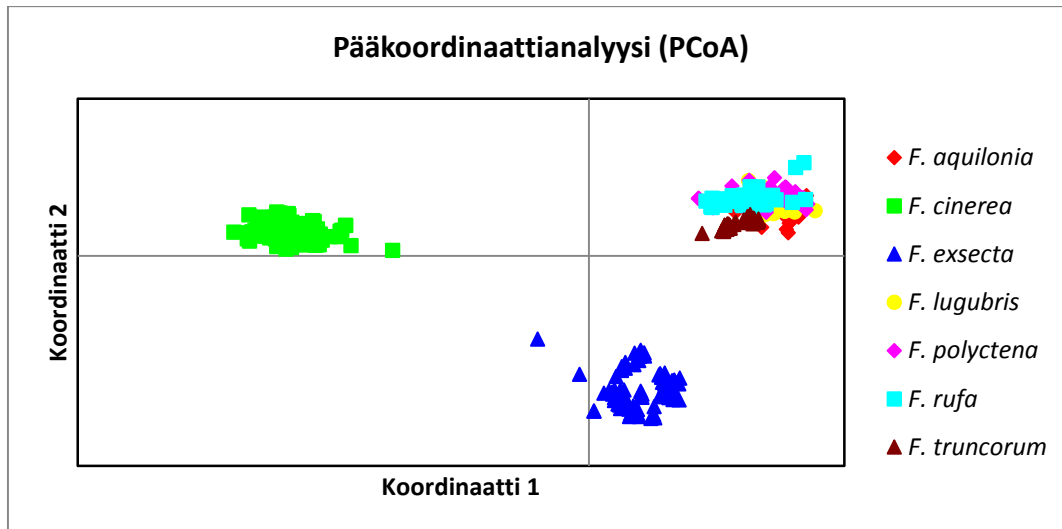
Kaikkien neljän geenin yhdistelmän perusteella tehdystä fylogeneettisestä puusta näkee, että *F. cinerea* ja *F. exsecta* eroavat selkeästi toisistaan ja muista lajeista (Kuva 4). *F. cinerea* on eronnut muista lajeista ensimmäisenä korkealla luotettavuudella (bootstrap-arvo 99%) (Kuva 4). *F. exsecta* on eronnut seuraavana samalla luotettavuudella. Puun mukaan myös *F. truncorum* eroaa muista *F. rufa* -ryhmän lajeista, mutta huomattavasti alhaisemmalla bootstrap-arvolla (32%). Muut *F. rufa* -ryhmän lajit eivät erotu puussa omiksi oksanhaaroikseen ja ovat tässä esitetty yhtenä haarana.



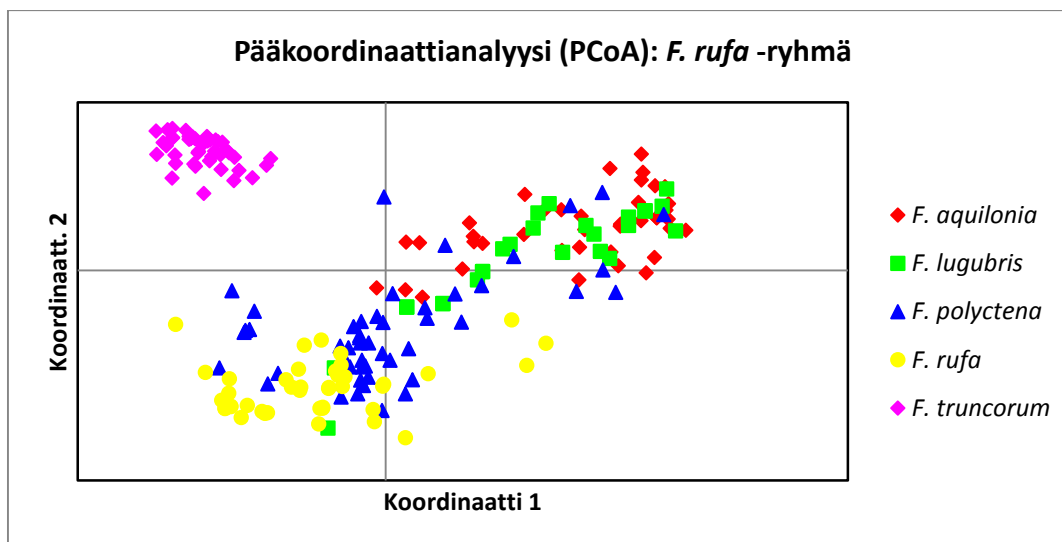
Kuva 4. *F. cinerean*, *F. exsectan* sekä *F. truncorumin* ja *F. rufa* -ryhmän muiden lajien fylogenia yhteen liitettyjen CSP1-, CSP7-, OBP1- ja OBP10-geenien perusteella. Puu tehtiin Mega:lla neighbor-joining -menetelmällä. Oksissa näkyy bootstrap-arvot 1000 uudelleenotannalla.

5.2.3. Pääkoordinaattianalyysi

Koska fylogeneettisestä puusta oli hankala erottaa *F. rufa* -ryhmän lajeja omina ryhminään, aineistoa tarkasteltiin vielä PCoA-analyysillä. Kaikkien yksilöiden välinen PCoA-analyysi, jossa on mukana kaikki geenit yhteen liitettynä, näyttää että *F. cinerean* ja *F. exsectan* yksilöt eroavat selvästi omiksi ryhmikseen ja *F. rufa* -ryhmän lajit muodostavat kolmannen ryhmän (Kuva 5a). PCoA-kuva, jossa on suurennettu vain *F. rufa* -ryhmän jäsenet, näyttää että vielä *F. truncorum* erottuu omaksi ryhmäkseen (Kuva 5b). Lisäksi *F. rufa* ja *F. aquilonia* erottuvat toisistaan, mutta muuten *F. rufa* -ryhmän lajit menevät koordinaatistossa päällekkäin.



Kuva 5a. Pääkoordinaattianalyysikuvaaja, jossa on mukana kaikkien tutkittavien muurahaisyksilöiden sekvenssit CSP1-, CSP7-, OBP1- ja OBP10-geeneistä. *F. cinerean*, *F. exsectan* ja *F. rufa* -ryhmän yksilöt on merkitty eri väreillä.



Kuva 5b. Pääkoordinaattianalyysikuvaaja, jossa on mukana *F. rufa* -ryhmän lajien yksilöiden sekvenssit CSP1-, CSP7-, OBP1- ja OBP10-geeneistä.

5.2.4. Nukleotididiversiteetti

Nukleotididiversiteetit laskettiin koko geeneille (sisältäen koodaavan ja ei-koodaavan alueen) *F. cinereassa*, *F. exsectassa*, *F. truncorumissa*, muissa *F. rufa* -ryhmän lajeissa ja kaikissa lajeissa yhteensä. Lajien tulokset vaihtelevat välillä 0,00003–0,0044 (Taulukko 6), joista korkeimpia ovat *F. cinerean* ja yhdistetyn *F. rufa* -ryhmän nukleotididiversiteetit. Koko aineiston nukleotididiversiteetit vaihtelevat välillä 0,00561–0,00857.

Taulukko 6. CSP1-, CSP7-, OBP1- ja OBP10-geenien nukleotididiversiteetit *F. cinereassa*, *F. exsectassa*, *F. truncorumissa*, muissa *F. rufa* -ryhmän lajeissa ja kaikissa lajeissa yhteensä.

Geeni	<i>F. cinerea</i>	<i>F. exsecta</i>	<i>F. truncorum</i>	<i>F. rufa</i> -ryhmä	Kaikki lajit
CSP1	0,00440	0,00091	0,00033	0,00233	0,00679
CSP7	0,00406	0,00170	0,00095	0,00114	0,00857
OBP1	0,00194	0,00197	0,00003	0,00305	0,00561
OBP10	0,00373	0,00159	0,00116	0,00215	0,00752

5.3. Sekvenssimuuntelu populaatioissa

5.3.1. F_{ST}

Tajiman D ja Fu ja Li -evoluutioanalyysien oletuksena on panmiktinen populaatio, jossa ei ilmene populaatiorakennetta. Tästä syystä populaatioiden erilaistumista mitattiin F_{ST} -arvoilla. *F. cinerean* monogynisten (CinKSK ja CinLI) ja polygynisten (CinKU ja CinTA) populaatioiden väliset parittaiset F_{ST} -arvot vaihtelevat välillä 0,02–0,16 (Taulukko 7a). Osa niistä on merkitseviä merkitsevyystasolla 0,05 ja CinTA/CinKU ja CinTA/CinLI populaatioiden väliset arvot ovat merkitseviä jopa merkitsevyystasolla 0,01. Tämän tuloksen vahvistaa aiemmat neljään polymorfiseen mikrosatelliittilokukseen perustuvat merkitsevät F_{ST} -arvot kaikkien neljän populaation välillä (Sirviö ym. 2010). *F. cinerean* KSK (mono), KU (poly), LI (mono) ja TA (poly) -populaatioita ei siis voida yhdistää yhdeksi populaatioksi evoluutioanalyysijä varten.

Taulukko 7a. Parittaiset F_{ST} -arvot ja merkitsevyydet *F. cinerean* monogynisten ja polygynisten populaatioiden välillä. Merkitsevyys: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Populaatio	CinKSK (mono)	CinKU (poly)	CinLI (mono)	CinTA (poly)
CinKSK (mono)	0.00000			
CinKU (poly)	0.02057	0.00000		
CinLI (mono)	0.08251 *	0.08090 *	0.00000	
CinTA (poly)	0.03057	0.07729 **	0.16077 **	0.00000

F. exsectan monogyniset populaatiot (ExA1-ExA8) yhdistettiin yhdeksi isoksi monogyniseksi populaatioksi, koska jokaisesta osapopulaatiosta oli sekvensoitu vain kolme yksilöä, eikä F_{ST} -arvoja ole mielekäästä laskea näin pienten otosten välillä. Yhdistämistä tukee myös se, että monogynisiin pesiin liittyy pitkä dispersaali ja kaikki osapopulaatiot ovat Oulun seudulta, joten vahvaa

populaatiorakennetta ei odoteta. Evoluutioanalyysien tuloksia tulkitessa täytyy kuitenkin muistaa aineiston mahdollinen heterogeenisyys. ExA1 - ExA8 populaatioista yhdistetyn suuren monogynisen populaation ja polygynisen ExA9 populaation välinen F_{ST} -arvo 0,02876 ei ole merkitsevä (Taulukko 7b). Kaikki *F. exsectan* monogyniset populaatiot voidaan siis yhdistää polygynisen populaation kanssa evoluutioanalyysiin.

Populaatio	Ex_mono	Ex_poly
Ex_mono	0,00000	
Ex_poly	0,02876	0,00000

Taulukko 7b. Parittaiset F_{ST} -arvot ja merkitsevyydet *F. exsectan* monogynisten ja polygynisten populaatioiden välillä. Merkitsevyys: * $P < 0,05$.

F. truncorum monogyniset (FTTvFsk, FTTvJsk ja FTTvMsk) populaatiot yhdistettiin yhdeksi isommaksi monogyniseksi populaatioksi samoin edellytyksin kuin edellä, mutta pitäen mielessä aineiston mahdollinen heterogeenisyys. Parittaisiin F_{ST} -arvoihin otettiin monogynisen populaation lisäksi mukaan polygyninen FTTv-populaatio ja mahdollisesti polygyninen FTKop-populaatio. Näiden kolmen populaation väliset parittaiset F_{ST} -arvot vaihtelevat välillä 0,14–0,22 (Taulukko 7c). Näistä ovat merkitseviä monogynisen ja polygynisen FTTv-populaatioiden välinen sekä polygynisen FTTv ja mahdollisesti polygynisen FTKop -populaatioiden välinen F_{ST} -arvo. Evoluutioanalyysit tehtiin erikseen monogynisille ja polygynisille populaatioille.

Populaatio	FTTv_mono	FTTv_poly	FTKop
FTTv_mono	0,00000		
FTTv_poly	0,20742 *	0,00000	
FTKop	0,14737	0,22294 *	0,00000

Taulukko 7c. Parittaiset F_{ST} -arvot ja merkitsevyydet *F. truncorum* monogynisten ja polygynisten populaatioiden välillä. Merkitsevyys: * $P < 0,05$.

5.3.2. Monogyniset ja polygyniset populaatiot

F. cinerean, *F. exsectan* ja *F. truncorum* mono- ja polygynisistä sekvenssiaineistoista etsittiin näiden kahden sosiaalisen muodon välisiä systemaattisia eroja. Mono- ja polygynisten näytteiden välillä ei ollut fiksoituneita eroja. Populaatioiden välillä oli kuitenkin hieman eroa SNP:ien esiintymisessä. Taulukkoon 8 merkittiin vain ne geenit, joissa on useampi kuin yksi SNP, joka esiintyy vain polygynisissä populaatioissa. Suurin osa näistä kohdista sijaitsi introneissa. Vain CSP1:ssä ja OBP1:ssä oli yksi tai kaksi vastaavaa kohtaa eksoneissa. *F. cinerean* osalta merkittiin lisäksi, esiintyykö SNP molemmissa vai pelkästään toisessa polygynisessä populaatiossa. Vain yksi

SNP:istä esiintyy molemmissa populaatioissa. *F. exsecta* aineistossa oli vain yksi polygyninen populaatio.

Taulukko 8. *F. cinerea* ja *F. exsecta* CSP1- ja CSP7-geenien eksoni- ja intronialueiden SNP:it, jotka ovat läsnä vain polygynisissä populaatioissa sekä niiden alleelifrekvenssit polygynisissä aineistoissa. N tarkoittaa polygynisten sekvenssien kokonaislukumäärää. Muutossarake kertoo SNP:ien sijainnit ja laadut. *F. cinerea* osalta on merkitty, onko SNP läsnä molemmissa vai pelkästään toisessa polygynisessä populaatiossa.

Geeni	Laji	N	Alue	Muutos	Alleelifrek.	Pop.
CSP1	<i>F. exsecta</i>	30	Eksoni	68T>C	0,13	
				868C>T	0,13	
			Introni	256G>A	0,10	
				265C>T	0,10	
				292G>A	0,13	
CSP7	<i>F. cinerea</i>	40	Introni	587T>C	0,10	KU
				742C>A	0,10	KU
				763C>T	0,13	KU, TA
				821A>C	0,10	KU
				957G>A	0,05	TA
				967G>T	0,05	TA
	<i>F. exsecta</i>	12	Introni	398T>A	0,33	
				720C>T	0,17	
OBP1	<i>F. exsecta</i>	42	Eksoni	1217G>A	0,10	
			Introni	559A>G	0,10	

5.4. Luonnonvalinta

5.4.1. CSP1

CSP1-geenille ei ole merkitseviä tuloksia McDonald-Kreitman -testistä millään lajilla (Taulukko 9). MK-testin mukaan ei siis voida sanoa, että aineisto poikkeaisi neutraalista evoluutiomallista. *F. pratensis* ei voida määrittää NI-arvoa vähäisen polymorfismin vuoksi (*F. pratensis* sekvenssien välillä ei ole synonyymisiä muutoksia). MK-tulostaulukko, jossa näkyy synonyymisten ja ei-synonyymisten polymorfismien ja lajien välisten fiksoituneiden erojen lukumäärät jokaiselle lajille ja geenille, on liitteenä (Liite 1).

Taulukko 9. McDonald-Kreitman -testin alfa ja NI -tulokset sekä merkitsevyys CSP1-geenille. N-sarakkeessa on testissä käytettyjen sekvenssien määrä. Merkitsevyys: * 0.01<P<0.05; ** 0.001<P<0.01; *** P<0.001.

Laji	N	Alfa	NI	Fisherin tarkka testi (p-arvo)	G-testi (p-arvo)
<i>F. aquilonia</i>	34	1,000	0,000	1,00000	-
<i>F. cinerea</i>	101	-1,333	2,333	0,57786	0,41826
<i>F. exsecta</i>	63	1,000	0,000	1,00000	-
<i>F. lugubris</i>	26	1,000	0,000	1,00000	-
<i>F. polycтена</i>	46	-0,438	1,438	1,00000	0,78253
<i>F. pratensis</i>	16	-	-	0,08021	-
<i>F. rufa</i>	42	1,000	0,000	1,00000	-
<i>F. truncorum</i>	48	-1,875	2,875	0,47727	0,48088

Tajiman D ja Fu ja Li -testit tehtiin *F. cinerean*, *F. exsectan* ja *F. truncorumin* suurimmille populaatioille. *F. cinerean* polygynisen KU-populaation Tajiman D on positiivinen ja merkitsevä, mikä viittaa tasapainottavaan valintaan (Taulukko 10). Muilla lajeilla ei ole merkitseviä Tajiman D tai Fu ja Li -testituloksia CSP1-geenistä. *F. truncorumin* polygynisestä populaatiosta ei ole ollenkaan tuloksia vähäisen polymorfismin vuoksi.

Taulukko 10. Tajiman D, Fu & Li D ja Fu & Li F -testien tulokset *F. cinerean*, *F. exsectan* ja *F. truncorumin* populaatioiden CSP1-geenille. N-sarakkeessa on testissä käytettyjen sekvenssien lukumäärä. Merkitsevät tulokset on merkitty # (P<0.10), * (P<0.05) tai ** (P<0.02). Tasapainottavaan valintaan viittaava tulos on merkitty vihreällä.

Laji/Populaatio	N	Tajiman D	Fu & Li D	Fu & Li F
<i>F. cinerea</i> KSK	26	0,4783	0,61206	0,81037
<i>F. cinerea</i> KU	28	2,13452 *	0,44485	1,05692
<i>F. cinerea</i> LI	18	0,25029	0,38879	0,4139
<i>F. cinerea</i> TA	14	0,27827	-0,08204	0,02058
<i>F. exsecta</i> Oulu	62	-1,23933	0,99146	0,40634
<i>F. truncorum</i> mono	24	-0,68111	0,60475	0,30632
<i>F. truncorum</i> poly	12	-	-	-

Koko geenin lisäksi Tajiman D ja Fu ja Li -testit tehtiin sliding window -vaihtoehdolla 100 nt pituisille sekvenssipätkille. Näiden tulosten mukaan *F. cinerean* KSK ja KU -populaatioilla on CSP1-geenissä mahdollista tasapainottavaa valintaa geenin ainoassa intronissa (Taulukko 11).

Taulukko 11. Geenin CSP1 Tajiman D ja Fu ja Li -analyysien merkitsevät sliding window -tulokset. N kertoo sekvenssien lukumäärän. Vihreä väri tarkoittaa merkitsevää tasapainottavaa valintaa. Merkitsevyys: # ($P < 0.10$), * ($P < 0.05$) tai ** ($P < 0.02$).

Populaatio	N	Alue (nt)	Eksoni/introni	Testisuure	Testi
<i>F. cinerea</i> KSK	26	199-493	eksoni1-introni1	1,7939 #	Tajima
<i>F. cinerea</i> KU	28	196-490	eksoni1-introni1	2,1078 *	Tajima

MFDM-testi tehtiin kolmelle lajille: *F. cinerealle*, *F. exsectalle* ja *F. truncorumille*. *F. exsectan* aineistosta löytyi mahdollista positiivista valintaa CSP1-geenin ensimmäisestä intronista (Taulukko 12). Tulos ei kuitenkaan ole merkitsevä rekombinaatiokorjauksen jälkeen.

Taulukko 12. MFDM-testin tulokset *F. cinerean*, *F. exsectan* ja *F. truncorumin* CSP1-geenille. N-sarakkeessa on testissä käytettyjen sekvenssien määrä. Jos p-arvo on pienempi kuin merkitsevyysraja (0,05), geeni voi olla positiivisen valinnan alainen. Alue-sarakkeessa on positiivisen valinnan alainen alue. Jos geenistä löytyy epätasapainoinen puu, MFDM ennustaa mahdollisten rekombinaatiotapahtumien (Rm) lukumäärän ja muuttaa tuloksen merkitsevyysrajaa sen mukaisesti.

Laji	N	Alue	p-arvo	Rm	Merkitsevyysraja
<i>F. cinerea</i>	73	-	0,972222	-	0,05
<i>F. exsecta</i>	55	introni	0,037037	9	0,005
<i>F. truncorum</i>	44	-	1	-	0,05

Transkriptiofaktoreiden sitoutumiskohdat ennustettiin introneille, joissa on merkkejä valinnasta Tajiman D:n tai Fu ja Li -testin perusteella. Tajiman D ja Fu ja Li -testit tehtiin vain *F. cinerealle*, *F. exsectalle* ja *F. truncorumille*, joista oli suurimmat sekvenssiaineistot. CSP1-geenin intronissa on neljän transkriptiofaktorin mahdollisia sitoutumiskohtia: Ftz, Hb, Eve ja Zen-1. Osalla *F. cinerean* näytteistä on SNP yhdessä Hb-transkriptiofaktorin sitoutumiskohdassa, mutta kyseinen SNP ei osu *F. cinereassa* tunnistetun mahdollisen tasapainottavan valinnan alueen alueelle.

5.4.2. CSP7

CSP7-geenin kohdalla *F. cinerean* McDonald-Kreitman -testin tulos on merkitsevä molempien merkitsevyystestien perusteella (Taulukko 13). Ni:n arvo on yli 1, joka viittaa mahdolliseen puhdistavaan valintaan. *F. aquilonian*, *F. exsectan* ja *F. lugubriksen* tulokset puuttuvat, koska niiden näytteissä ei ole lajin sisäistä polymorfismia CSP7-geenissä.

Taulukko 13. McDonald-Kreitman -testin alfa ja NI -tulokset sekä merkitsevyys CSP7-geenille. N-sarakkeessa on testissä käytettyjen sekvenssien määrä. Merkitsevyys: * 0.01<P<0.05; ** 0.001<P<0.01; *** P<0.001.

Laji	N	Alfa	NI	Fisherin tarkka testi (p-arvo)	G-testi (p-arvo)
<i>F. aquilonia</i>	28	-	-	-	-
<i>F. cinerea</i>	106	-5,133	6,133	0,030521 *	0,01308 *
<i>F. exsecta</i>	32	-	-	-	-
<i>F. lugubris</i>	17	-	-	-	-
<i>F. polycytena</i>	48	1,000	0,000	0,548263	-
<i>F. rufa</i>	32	0,212	0,788	1,000000	0,84128
<i>F. truncorum</i>	22	1,000	0,000	1,000000	-

CSP7-geenissä *F. exsectan* Oulusta kerätyn populaation Tajima ja Fu ja Li -tulokset ovat merkitseviä tasolla $p < 0,05$ (Taulukko 14). Tulokset ovat negatiivisia, mikä viittaa viimeaikaiseen positiiviseen tai puhdistavaan valintaan tai populaation koon nopeaan kasvuun. *F. truncorumin* monogynisen populaation Tajiman D -tulos on merkitsevä tasolla $p < 0,10$. Positiivinen D-arvo viittaa viimeaikaiseen tasapainottavaan valintaan tai populaation koon pienentymiseen. *F. truncorumin* polygynisen populaation aineisto oli liian pieni analyysiin.

Taulukko 14. Tajiman D, Fu & Li D ja Fu & Li F -testien tulokset *F. cinerean*, *F. exsectan* ja *F. truncorumin* populaatioiden CSP7-geenille. N-sarakkeessa on testissä käytettyjen sekvenssien määrä. Merkitsevät tulokset on merkitty # ($P<0.10$), * ($P<0.05$) tai ** ($P<0.02$). Puhdistava tai positiivinen valinta on merkitty punaisella ja tasapainottava valinta vihreällä.

Laji/Populaatio	N	Tajiman D	Fu & Li D	Fu & Li F
<i>F. cinerea KSK</i>	22	-1,54163	0,28059	0,30516
<i>F. cinerea KU</i>	24	-0,10769	0,85933	0,89423
<i>F. cinerea LI</i>	22	-0,28351	0,71813	0,39052
<i>F. cinerea TA</i>	22	-1,19166	0,52765	0,19472
<i>F. exsecta Oulu</i>	24	-1,9913 *	-2,44664 *	-2,73414 *
<i>F. truncorum mono</i>	14	1,74339 #	0,67726	1,01617
<i>F. truncorum poly</i>	2	-	-	-

Tajiman D:n ja Fu ja Li -testin sliding window -versiot tunnistivat CSP7-geenistä mahdollisia valinnan alueita 100 nt:n ikkunoilla. Osa testien tunnistamista positiivisen tai puhdistavan tai tasapainottavan valinnan alueista osui CSP7:n introniin (Taulukko 15).

Taulukko 15. Geenin CSP7 Tajiman D ja Fu ja Li -analyysien merkitsevät sliding window -tulokset. N kertoo sekvenssien lukumäärän. Vihreä väri tarkoittaa merkitsevää tasapainottavaa valintaa ja punainen merkitsevää positiivista tai puhdistavaa valintaa. Merkitsevyys: # (P<0.10), * (P<0.05) tai ** (P<0.02).

Populaatio	N	Alue (nt)	Eksoni/introni	Testisuure	Testi
<i>F. cinerea</i> KSK	22	661-767	introni1	-1,6671 #	Tajima
<i>F. cinerea</i> KSK	22	1080-1188	eksoni2	1,7830 *	Fu&Li
<i>F. cinerea</i> LI	22	460-612	introni1	1,9175 #	Tajima
<i>F. cinerea</i> TA	22	658-815	introni1	-1,7294 #	Tajima
<i>F. exsecta</i> Oulu	24	167-343	eksoni1-introni1	-1,8838 *	Tajima
<i>F. exsecta</i> Oulu	24	157-365	eksoni1-introni1	-2,2652 #	Fu&Li
<i>F. truncorum</i> mono	14	298-372	introni1	1,7434 #	Tajima

MFDM testi havaitsee CSP7-geenin ensimmäisestä intronissa mahdollista positiivista valintaa *F. cinerean* ja *F. exsectan* aineistoista merkitsevyydellä $p < 0,10$ (Taulukko 16).

Rekombinaatiokorjauksen jälkeen kumpikaan tuloksista ei ole merkitsevä. Lisäksi *F. cinerean* merkitsevä tulos voi johtua multiple-hit:istä. Muutamalla *F. cinerean* yksilöllä on kyseisessä kohdassa T-nukleotidi toisessa tai molemmissa alleeleissa, ja kaikilla muilla yksilöillä on C.

Ulkoryhmänä on käytetty *F. exsectaa*, jolla on kyseisessä kohdassa T. *F. cinerean* T muutamassa yksilössä saattaa johtua myös uudesta mutaatiosta, eikä siitä että C olisi yleistynyt positiivisen valinnan kautta.

Taulukko 16. MFDM-testin tulokset *F. cinerean*, *F. exsectan* ja *F. truncorum*in CSP7-geenille. N-sarakkeessa on testissä käytettyjen sekvenssien määrä. Jos p-arvo on pienempi kuin merkitsevyysraja (0,10), geeni voi olla positiivisen valinnan alainen. Alue-sarakkeessa on positiivisen valinnan alainen alue. Jos geenistä löytyy epätasapainoinen puu, MFDM ennustaa mahdollisten rekombinaatiotapahtumien lukumäärän ja muuttaa tuloksen merkitsevyysrajaa sen mukaisesti. * merkitty tulos voi johtua multiple-hit:istä.

Laji	N	Alue	p-arvo	Rm	Merkitsevyysraja
<i>F. cinerea</i>	81	introni	0,075*	30	0,003226
<i>F. exsecta</i>	24	introni	0,086957	7	0,0125
<i>F. truncorum</i>	18	-	-	-	0,05

CSP7-geenin introniin ennustettiin Hb:n, Ftz:n, Tll:n, Prd:n, Dfd:n, DSXF:n, DSXM:n ja B-faktorin sitoutumiskohtia. Hb:n ja Ftz:n sitoutumiskohtien määrä erosi *F. cinerean* ja *F. exsectan* ja muiden lajien välillä, johtuen sekä indeleistä että SNP:eistä. Vain yksi SNP osui tunnistettuun positiivisen tai puhdistavan valinnan alueeseen. Kohdassa on mahdollinen Tll-faktorin sitoutumiskohta. *F. rufa* -ryhmän lajeilla on SNP-kohdassa aina C-nukleotidi, *F. exsectalla* aina T ja *F. cinerealla* yleensä C,

mutta muutamalla näytteellä T. Kyseessä on sama kohta, jonka MFDM-testi tunnistaa mahdollisesti positiivisen valinnan alaiseksi.

5.4.3. OBP1

McDonald-Kreitman -testin tulokset OBP1-geenissä *F. aquilonian* ja *F. lugubriksen* osalta ovat merkitseviä, *F. aquilonia* Fisherin tarkan testin mukaan ja *F. lugubris* G-testin mukaan (Taulukko 17). Molempien NI on alle 1, joka viittaa mahdolliseen positiiviseen valintaan. *F. truncorum* tuloksia ei voida määrittää, koska sen aineistossa ei ole polymorfismia OBP1-geenissä.

Taulukko 17. McDonald-Kreitman -testin alfa ja NI -tulokset sekä merkitsevyys OBP1-geenille. N-sarakkeessa on testissä käytettyjen sekvenssien määrä. Merkitsevyys: * 0.01<P<0.05; ** 0.001<P<0.01; *** P<0.001.

Laji	N	Alfa	NI	Fisherin tarkka testi (p-arvo)	G-testi (p-arvo)
<i>F. aquilonia</i>	40	1,000	0,000	0,040704*	-
<i>F. cinerea</i>	138	0,733	0,267	1,000000	0,82997
<i>F. exsecta</i>	101	0,267	0,733	0,321549	0,21543
<i>F. lugubris</i>	28	0,853	0,147	0,087335	0,04914*
<i>F. polyclena</i>	50	1,000	0,000	0,092727	-
<i>F. rufa</i>	38	1,000	0,000	0,168440	-
<i>F. truncorum</i>	48	-	-	-	-

F. exsectan Oulu-populaatiolla on merkitsevä positiivinen Fu ja Li -tulos (Taulukko 18), joka viittaa mahdolliseen tasapainottavaan valintaan. *F. truncorum* populaatioista ei ole ollenkaan tuloksia vähäisen polymorfismin vuoksi.

Taulukko 18. Tajiman D, Fu & Li D ja Fu & Li F -testien tulokset *F. cinerean*, *F. exsectan* ja *F. truncorum* populaatioiden OBP1-geenille. N-sarakkeessa on testissä käytettyjen sekvenssien määrä. Merkitsevät tulokset on merkitty # (P<0.10), * (P<0.05) tai ** (P<0.02). Tasapainottava valinta on merkitty vihreällä.

Laji/Populaatio	N	Tajiman D	Fu & Li D	Fu & Li F
<i>F. cinerea</i> KSK	31	0,36308	-0,31508	-0,29642
<i>F. cinerea</i> KU	38	0,26698	0,92501	0,73115
<i>F. cinerea</i> LI	22	1,70562	0,84065	1,08773
<i>F. cinerea</i> TA	26	-0,29232	0,97343	0,9008
<i>F. exsecta</i> Oulu	90	-0,50176	1,51454 *	0,79412
<i>F. truncorum</i> mono	24	-	-	-
<i>F. truncorum</i> poly	12	-	-	-

Tajiman D ja Fu ja Li sliding window -testien mukaan *F. cinerea*lla on OBP1-geenissä mahdollista positiivista tai puhdistavaa valintaa intronissa 4 ja mahdollista tasapainottavaa valintaa intronissa 1 (Taulukko 19).

Taulukko 19. Geenin OBP1 Tajiman D ja Fu ja Li -analyysien merkitsevät sliding window -tulokset. N kertoo sekvenssien lukumäärän. Vihreä väri tarkoittaa merkitsevää tasapainottavaa valintaa ja punainen merkitsevää positiivista tai puhdistavaa valintaa. Merkitsevyys: # ($P < 0.10$), * ($P < 0.05$) tai ** ($P < 0.02$).

Populaatio	N	Alue (nt)	Eksoni/introni	Testisuure	Testi
<i>F. cinerea</i> KSK	25	1039-1163	introni4	-1,7333 #	Tajima
<i>F. cinerea</i> KSK	25	1125-1309	introni4-eksoni5	-2,7174 *	Fu&Li
<i>F. cinerea</i> LI	22	105-211	introni1	1,8281 #	Tajima

MFDM-testin perusteella OBP1-geenissä ei ole merkkejä positiivisesta luonnonvalinnasta, koska tulokset eivät ole merkitseviä (Taulukko 20).

Taulukko 20. MFDM-testin tulokset *F. cinerean*, *F. exsectan* ja *F. truncorum*in OBP1-geenille. N-sarakkeessa on testissä käytettyjen sekvenssien määrä. Kaikkien lajien p-arvot ovat merkitsevyysrajaa (0,05) suurempia, joten geenissä ei ole merkkejä positiivisesta valinnasta.

Laji	N	Alue	p-arvo	Rm	Merkitsevyysraja
<i>F. cinerea</i>	130	-	0,666667	-	0,05
<i>F. exsecta</i>	102	-	0,316832	-	0,05
<i>F. truncorum</i>	44	-	1	-	0,05

Tajiman D ja Fu ja Li -testien sliding window -versiot tunnistivat mahdollista valintaa *F. cinerean* OBP1-geenin introneissa 1 ja 4. OBP1:n ensimmäisessä intronissa on Mad:in, Prd:n ja Ftz:n mahdollisia sitoutumiskohtia. Neljännessä intronissa on Mad:in ja Ftz:n sitoutumiskohtia. Niissä ei ole lajien välisiä tai sisäisiä eroja. Toisessa intronissa on Prd:n, Ftz:n ja Hb:n sitoutumiskohtia. *F. cinerea* ja *F. exsecta* eroavat Prd:n osalta muista lajeista, mutta kyseistä aluetta ei tunnistettu valinnan alaiseksi kummallakaan lajilla.

5.4.4. OBP10

OBP10-geenille ei ole merkitseviä tuloksia McDonald-Kreitman -testistä (Taulukko 21).

Taulukko 21. McDonald-Kreitman -testin alfa ja NI -tulokset sekä merkitsevyys OBP10-geenille. N-sarakkeessa on testissä käytettyjen sekvenssien määrä. Merkitsevyys: * 0.01<P<0.05; ** 0.001<P<0.01; *** P<0.001.

Laji	N	Alfa	NI	Fisherin tarkka testi (p-arvo)	G-testi (p-arvo)
<i>F. aquilonia</i>	35	-3,071	4,071	0,295514	0,16089
<i>F. cinerea</i>	88	-3,750	4,750	0,28603	0,12387
<i>F. exsecta</i>	96	1,000	0,000	0,569231	-
<i>F. lugubris</i>	22	1,000	2,000	0,641919	0,47194
<i>F. polyclena</i>	42	-3,500	4,500	0,283268	0,14539
<i>F. pratensis</i>	10	1,000	0,000	1,000000	-
<i>F. rufa</i>	46	-4,429	5,429	0,220033	0,17789

F. exsectan Oulu-populaation Tajiman D -tulos on negatiivinen ja merkitsevä tasolla $p > 0,10$ (Taulukko 22). Tulos viittaa mahdolliseen puhdistavaan tai adaptiiviseen valintaan. OBP10-geenin kohdalla ei ole merkitseviä Tajiman D tai Fu ja Li -tuloksia muille lajeille (Taulukko 22).

Taulukko 22. Tajiman D, Fu & Li D ja Fu & Li F -testien tulokset *F. cinerean*, *F. exsectan* ja *F. truncorum* populaatioiden OBP10-geenille. N-sarakkeessa on testissä käytettyjen sekvenssien määrä. Merkitsevät tulokset on merkitty # (P<0.10), * (P<0.05) tai ** (P<0.02). Puhdistava tai positiivinen valinta on merkitty punaisella.

Laji/Populaatio	N	Tajiman D	Fu & Li D	Fu & Li F
<i>F. cinerea</i> KSK	22	-1,15235	-0,91985	-1,26298
<i>F. cinerea</i> KU	32	0,46695	0,41098	0,45861
<i>F. cinerea</i> LI	14	-0,32092	-0,51913	-0,60807
<i>F. cinerea</i> TA	12	-1,17901	-0,48275	-0,66813
<i>F. exsecta</i> Oulu	88	-1,77928 #	-1,0777	-1,45324
<i>F. truncorum</i> mono	24	-0,18977	0,98473	0,72527
<i>F. truncorum</i> poly	12	-0,46008	-0,7119	-0,77038

Tajiman D ja Fu ja Li sliding window -testien perusteella OBP10-geenissä on mahdollista positiivista tai puhdistavaa valintaa introneissa 1 (*F. cinerea*), 3 (*F. exsecta*) ja 4 (*F. cinerea* ja *F. exsecta*) (Taulukko 23).

Taulukko 23. Geenin OBP10 Tajiman D ja Fu ja Li -analyysien merkitsevät sliding window -tulokset. N kertoo sekvenssien lukumäärän. Punainen väri tarkoittaa merkitsevää positiivista tai puhdistavaa valintaa.

Merkitsevyys: # (P<0.10), * (P<0.05) tai ** (P<0.02).

Populaatio	N	Alue (nt)	Eksoni/introni	Testisuure	Testi
<i>F. cinerea</i> KSK	16	488-748	introni1	-2,0655 #	Fu&Li
<i>F. cinerea</i> LI	14	1571-1731	introni4	-2,1764 #	Fu&Li
<i>F. cinerea</i> LI	8	275-452	introni1	-1,9422 #	Fu&Li
<i>F. exsecta</i> Oulu	76	1106-1205	introni3	-1,6842 #	Tajima
<i>F. exsecta</i> Oulu	88	1122-1254	introni3	-2,0122 #	Fu&Li
<i>F. exsecta</i> Oulu	88	1640-1875	introni4	-2,0122 #	Fu&Li

OBP10-geenille ei tullut MFDM-tuloksia ollenkaan (Taulukko 24), mikä saattaa johtua sekvenssien puuttuvien alueiden suuresta määrästä.

Laji	N	Alue	p-arvo	Rm	Merkitsevyysraja
<i>F. cinerea</i>	94	-	-	-	0,05
<i>F. exsecta</i>	98	-	-	-	0,05
<i>F. truncorum</i>	48	-	-	-	0,05

Taulukko 24. MFDM-testin tulokset

F. cinerean, *F. exsectan* ja *F. truncorum* CSP1-geenille. N-sarakkeessa on testissä käytettyjen sekvenssien määrä.

Tajiman D ja Fu ja Li -testien sliding window -tuloksien mukaan joko *F. cinerealla* tai *F. exsectalla* on mahdollista valintaa OBP10:n introneissa 1, 3 ja 4. OBP10:n ensimmäisessä intronissa on Hb:n, Ftz:n, Eve:n, Zen-1:n ja Dfd:n sitoutumiskohtia. *F. exsecta* ja *F. polycytena* eroavat muista lajeista Ftz:n osalta johtuen SNP:eistä ja indeleistä. *F. exsectalla* ei ole kuitenkaan tunnistettu valintaa ensimmäisessä intronissa. *F. exsectalla* on lajin sisäistä muuntelua yhdessä Ftz:n sitoutumiskohdassa intronissa 2, mutta alueella ei havaittu valintaa. Neljännessä intronissa on Ftz:n, Dfd:n ja Prd:n sitoutumiskohtia. Yhdessä Ftz:n sitoutumiskohdan sekvenssissä on muuntelua *F. cinerean* sisällä ja *F. cinerealla* on tunnistettu mahdollista positiivista tai puhdistavaa valintaa kyseisellä alueella.

5.5. Rekombinaatio

SBP-rekombinaatiotesti tehtiin sekvenssirinnastuksille, jotka sisälsivät sekä koodaavan että ei-koodaavan alueen. Jokaiselle geenille ja lajille käytettiin kahta aineistoa. Ensimmäisessä aineistossa puuttuvat nukleotidit on ennustettu lajin sen kohdan yleisimmän nukleotidin mukaan. Toisessa aineistossa puuttuvat nukleotidit on jätetty puuttuviksi. SBP-rekombinaatiotestin perusteella vain CSP1-geenissä ei ole todisteita rekombinaatiosta (Taulukko 25). CSP7-geenissä on todisteita rekombinaatiosta *F. cinerea* ensimmäisen aineiston mukaan. OBP1-geenissä on todisteita rekombinaatiosta *F. cinerea* ja *F. exsecta* myös ensimmäisen aineiston mukaan. OBP10-geenissä on todisteita rekombinaatiosta *F. cinerea* ja *F. truncorum* molempien aineistojen perusteella.

Taulukko 25. SBP-rekombinaatiotestin tulokset jokaiselle geenille lajeittain käyttäen kahta aineistoa. Ensimmäisessä aineistossa puuttuvat kohdat on ennustettu lajin yleisimmäksi nukleotidiksi (1) ja toisessa puuttuvat nukleotidit on merkitty aina N-kirjaimella (2).

		1)			2)		
Geeni	Laji	AIC	AICc	BIC	AIC	AICc	BIC
CSP1	<i>F. cinerea</i>	-	-	-	-	-	-
	<i>F. exsecta</i>	-	-	-	-	-	-
	<i>F. truncorum</i>	-	-	-	-	-	-
CSP7	<i>F. cinerea</i>	+	-	-	-	-	-
	<i>F. exsecta</i>	-	-	-	-	-	-
	<i>F. truncorum</i>	-	-	-	-	-	-
OBP1	<i>F. cinerea</i>	+	+	-	-	-	-
	<i>F. exsecta</i>	+	+	-	-	-	-
	<i>F. truncorum</i>	-	-	-	-	-	-
OBP10	<i>F. cinerea</i>	+	+	-	+	+	-
	<i>F. exsecta</i>	-	-	-	-	-	-
	<i>F. truncorum</i>	+	-	-	+	-	-

6. TULOSTEN TARKASTELU

6.1. Sekvenssimuuntelu *F. rufa* -ryhmän lajien välillä on vähäistä

Tässä työssä tutkittiin sekvenssimuuntelun määrää läheisten *Formica*-muurahaislajien välillä sekä lajin sisäistä muuntelua hajukommunikaation taustalla olevissa geeneissä. Tutkimuksessa lähdettiin oletuksesta, että aineisto sisälsi seitsemän toisistaan eroavaa muurahaislajia. Sekvenssivertailut kuitenkin osoittivat, että tutkittavissa geeneissä oli hyvin vähän muuntelua. Tämä johti lisätutkimuksiin, jotka osoittivat, etteivät kaikki *F. rufa* -ryhmän lajit eroa toisistaan tämän työn aineiston perusteella.

Aineistoa ja lajien eroavuutta toisistaan tarkasteltiin fiksoituneilla muutoksilla, fylogeniaalla ja PCoA-analyysillä. Kaikki menetelmät antavat suunnilleen samaa tietoa, fiksoituneet muutokset ja PCoA vain tarkemmin kuin fylogeneettinen puu. Fylogeneettinen puu ja PCoA-analyysi tehtiin kaikkien neljän geenin yhdistelmälle, fiksoituneet muutokset taas jokaiselle geenille erikseen. Fylogeneettisen puun perusteella *F. cinerea* ja *F. exsecta* eroavat selkeästi muista lajeista ja *F. cinerea* eroaa omaksi oksanhaarakseen ensimmäisenä, mikä vastaa aiempaa mitokondriaalisen sekvenssiaineiston antamaa tietoa (Goropashnaya ym. 2012). *F. rufa* -ryhmän lajeista puun perusteella erottuu omaksi ryhmäkseen vain *F. truncorum* ja sekin alhaisella luotettavuudella (n. 30 %). Useampi laji voi yhdistyä yhdeksi ryhmäksi fylogeneettisissä analyyseissä lajien yhteisen polymorfian vuoksi. Tällöin kaikki kyseiset lajit ovat saattaneet saada useamman alleelin yhteiseltä kantaisältäään ja kaikki alleelit ovat säilyneet lajeissa joko valinnan vaikutuksesta tai lajien hiljattain tapahtuneen eroamisen vuoksi. Tällaista on todettu esimerkiksi *Drosophila* lajeilla (Charlesworth ym. 2005).

PCoA-kuvaajan mukaan *F. cinerea* ja *F. exsecta* eroavat toisistaan yhtä paljon kuin ne eroavat *F. rufa* -ryhmästä ja *F. rufa* -ryhmän sisäinen muuntelu on yhtä suurta kuin *F. cinerean* tai *F. exsectan* sisäinen muuntelu. Tämän perusteella *F. cinerea* ja *F. exsecta* ovat hyviä lajeja, koska ne erottuvat muista lajeista selkeästi omiksi ryhmikseen. Vaikka *F. rufa* -ryhmän lajit ovat PCoA-kuvaajassa lähekkäin, *F. truncorum* eroaa muista ja lisäksi *F. rufa* ja *F. aquilonia* eroavat toisistaan. PCoA-kuvaajan perusteella *F. lugubris*, *F. polychtena* ja *F. aquilonia* eivät erotu toisistaan eikä niiden välillä havaita yhtäkään fiksoitunutta eroa yhdessäkään geenissä, vaikka tarkasteltaisiin myös introneita. Näitä lajeja ei siis voida erottaa toisistaan tässä työssä käytettyjen geenien perusteella, eivätkä nämä geenit voi toimia markkereina lähilajien välillä tai lajintunnistuksessa. Fiksoituneista

muutoksista näkee eri geenien antaman tiedon lajien eroista. *F. cinerea* ja *F. exsecta* eroavat muista lajeista ja toisistaan kaikkien geenien perusteella. *F. rufa* -ryhmän sisäisistä eroista eri geenit antavat eri tietoa. *F. truncorum* eroaa muista lajeista OBP1-geenin ja osittain CSP-geenien mukaan. *F. rufa* eroaa *F. aquiloniasta* ja *F. lugubriksesta* CSP7-geenin mukaan. *F. polyctena* ei eroa *F. truncorumia* lukuun ottamatta muista *F. rufa* -ryhmän lajeista minkään geenin osalta.

F. rufa -ryhmän lajien sekvenssien samankaltaisuus voi johtua siitä, että lajiutumista on kulunut niin vähän aikaa, ettei tässä työssä tutkittuihin geeneihin ole ehtinyt kehittyä eroja lajien välillä. Lajien arvioidaan eronneen toisistaan viimeisen 500 000 vuoden aikana (Goropashnaya ym. 2004). *F. rufa* -ryhmän lajien välillä on kuitenkin todettu eroja ainakin mitokondriaalisten DNA-sekvenssien perusteella (Goropashnaya ym. 2012). Mitokondrion genomisekvenssissä onkin yleensä enemmän variaatiota kuin tuman genomissa (Brown ym. 1979). Goropashnayan ym. (2012) tulosten mukaan *F. polyctena* ja *F. rufa* ja toisaalta *F. aquilonia* ja *F. lugubris* ovat lähinnä toisiaan. Tässä aineistossa tuota jakoa ei kuitenkaan näy.

Toinen vähäisen muuntelun selittävä vaihtoehto on, että lajien välillä tapahtuu vielä risteytymistä ja aineistossa on mukana lajien välisiä hybridejä. Lajien välinen geenivirta voi yhdenmukaistaa sekvenssit niin, ettei fiksoituneita nukleotidieroja pääse syntymään eivätkä lajit pääse erkaantumaan toisistaan (Kulmuni ym. 2010). Viitteitä *F. rufa* -ryhmän lajien risteytymisestä on löydetty aiemminkin mitokondriaalisesta DNA:sta (Goropashnaya ym. 2004). *F. rufan* ja *F. polyctenan* risteytyminen keskenään on yleistä ja *F. aquilonian*, *F. pratensiksen* ja *F. lugubriksen* välillä on ollut risteytymistapahtumia aiemmissa sukupolvissa (Goropashnaya ym. 2004).

6.2. Luonnonvalinta muurahaisten hajugeeneissä

Työssä tutkittiin luonnonvalintaa neljässä muurahaisten hajugeenissä. Hajugeenit ovat mielenkiintoinen kohde luonnonvalinnan mittaamiselle, koska niissä on havaittu muita geeniryhmiä yleisemmin positiivista valintaa (Whiteman ym. 2008, Forêt ym. 2006, Vieira ym. 2012, Kulmuni ym. 2013a, Kulmuni ym. 2013b) ja niillä voi olla rooli lajiutumisessa (Smadja ym. 2009) ja sopeutumisessa. Tutkitut geenit löytyvät ortologeina useilta Hymenoptera-hyönteislajeilta, eli lajeilla on siis sama geenikopio. Aiemmat tutkimukset näille ortologisille OBP- ja CSP-geeneille ovat perustuneet kaukaisten lajien välisiin vertailuihin, eikä niissä ole havaittu positiivista valintaa (Kulmuni ym. 2013a, McKenzie ym. 2014). Tätä tutkimusta vastaavia

populaatio- tai lajitestejä, joissa tutkitaan valintaa yksittäisissä lajeissa, ei ole kuitenkaan aiemmin tehty näille geeneille. Tällaiset testit voivat kertoa hiljattain tapahtuneesta valinnasta, jota lajien väliset vertailut eivät havaitse.

Tässä työssä tutkittujen muurahaisten CSP- ja OBP-geenien nukleotididiversiteetit ovat joko samaa luokkaa tai vielä alhaisempia kuin aiempien tutkimusten hajugeenien nukleotididiversiteetit.

Drosophila-kärpäsen OR-geenien nukleotididiversiteetit vaihtelevat välillä 0,0014 – 0,0122

(Rollmann ym. 2010) ja OBP-geenien nukleotididiversiteetit välillä 0,001 – 0,018 (Wang ym. 2007).

Tämän työn geenien nukleotididiversiteetit vaihtelivat välillä 0,00561 – 0,00857 koko aineistolle ja näiden CSP- ja OBP-geenien nukleotididiversiteetit ovat siis alhaisia.

6.2.1. CSP7-geenissä on eniten merkkejä luonnonvalinnasta

Tässä työssä käytettiin neljää erilaista testiä luonnonvalinnan tutkimiseen sekvenssiaineistosta. Evoluutiovoimien testien tulokset eivät ole yhdenmukaisia. Eniten merkkejä luonnonvalinnasta havaittiin CSP7-geenissä (Taulukko 26). *F. cinerean* ja *F. exsectan* aineistoissa havaittiin sekä mahdollista positiivista että puhdistavaa valintaa.

F. cinerean tulokset ovat luotettavimpia, koska sen aineisto on laajin, eikä aineistossa ole populaation piilorakennetta. MK-testin mukaan *F. cinerealla* on geenissä merkkejä puhdistavasta valinnasta ja MFDM-testin mukaan positiivisesta valinnasta. Rekombinaatiotapahtumat geenissä saattavat kuitenkin vaikuttaa tuloksiin. MK- ja MFDM-testien tulokset ovat näennäisesti ristiriidassa keskenään. MK-testi käyttää kuitenkin vain eksoneita ja MFDM:n havaitsemat positiivisen valinnan alaiset alueet sijaitsevat intronissa. Puhdistava ja positiivinen valinta voivat vaikuttaa geenin eri osiin. CSP7-geenin eksonit voivat olla puhdistavan valinnan alaisia ja introni positiivisen valinnan alaisena. *F. exsectan* aineisto on huomattavasti pienempi ja aineistossa on yhdistettyjä populaatioita. *F. exsectalla* on CSP7-geenissä MFDM-testin mukaan positiivista valintaa ja Tajiman D ja Fu ja Li -testien mukaan positiivista tai puhdistavaa valintaa. MFDM-testin mukaan tulos voi johtua rekombinaatiosta, mutta SBP-rekombinaatiotesti ei ennusta rekombinaatiota *F. exsectan* sekvenssiaineistosta. Tajiman D ja Fu ja Li -testien merkitsevät tulokset *F. exsectan* osalta saattavat viitata myös populaation piilorakenteeseen aineistossa monogynisten populaatioiden yhdistämisen vuoksi. Tajiman D:n mukaan *F. truncorumilla* on geenissä tasapainottavaa valintaa. Tämäkin saattaa johtua populaation piilorakenteesta.

Positiivinen valinta sopii CSP7:n funktioon pesätoverien tunnistamisessa (Ozaki ym. 2005) ja hypoteesiin siitä, että CSP7 kehittyisi koevoluutiossa kutikulaaristen hiilivetyjen kanssa. Positiivisen valinnan alaisten kandidaattigeenien tunnistus on kuitenkin vasta ensimmäinen vaihe sopeutumisen taustan ymmärtämisessä. Lisäksi pitää selvittää, miten muutokset vaikuttavat proteiinin rakenteeseen ja funktioon. Pitäisi tietää sijoittuuko valinnan alainen kohta esimerkiksi laskostuneen proteiinin sitoutumistaskuun tai geenin introneihin transkriptiofaktoreiden sitoutumiskohtiin tai aiheuttaako se proteiinin pinnan varausmuutoksia.

Taulukko 26. Mahdollinen valinta CSP1-, CSP7-, OBP1-, ja OBP10-geeneissä eri valinnan testien perusteella. Harmaa ruutu tarkoittaa, ettei kyseiselle geenille ole kyseisestä analyysistä nollahypoteesista merkitsevästi poikkeavia tuloksia tai että tulokset puuttuvat. Merkitsevät tulokset on merkitty lajin nimellä ja symbolilla: - mahdollinen negatiivinen valinta, + mahdollinen positiivinen valinta tai * mahdollinen balansoiva valinta. MFDM-tulokset eivät ole merkitseviä, jos rekombinaatiokorjaus otetaan huomioon.

Geeni	MK-testi	Tajiman D	Fu & Li	MFDM
CSP1		<i>F. cinerea</i> KU *		<i>F. exsecta</i> +
CSP7	<i>F. cinerea</i> -	<i>F. exsecta</i> +/- <i>F. truncorum</i> *	<i>F. exsecta</i> +/-	<i>F. cinerea</i> + <i>F. exsecta</i> +
OBP1	<i>F. aquilonia</i> + <i>F. lugubris</i> +		<i>F. exsecta</i> *	
OBP10		<i>F. exsecta</i> +/-		

CSP1-geeni ekspressoituu *C. biroi* -muurahaisella (McKenzie ym. 2014) ja mehiläisellä (Forêt ym. 2007) vahvasti juuri tuntosarvissa ja toimii siis luultavasti hajumolekyylien sitomisessa. CSP1-geenissä on merkkejä valinnasta vain Tajiman D ja MFDM-testien perusteella. *F. exsectan* MFDM tulos viittaa positiiviseen valintaan CSP1:n intronissa (Taulukko 26). Merkitsevä tulos voi kuitenkin johtua geenissä tapahtuneista rekombinaatiotapahtumista, eikä tulos ole merkitsevä rekombinaatiokorjauksen jälkeen. Toisaalta HyPhy:n SBP-rekombinaatiotestin perusteella geenissä ei ole rekombinaatiota. Tajiman D -testin mukaan *F. cinerean* KU-populaatiossa on merkkejä tasapainottavasta valinnasta. Valinnan testien tulokset muilla lajeilla eivät anna selkeitä viitteitä siitä, että CSP1 olisi luonnonvalinnan vaikutuksen alaisena. Toisaalta *F. exsectan* ja *F. truncorum*in CSP1-aineistot ovat huomattavasti pienemmät verrattuna muihin geeneihin, mikä voi vaikuttaa analyysituloksiin.

OBP1-geeni ekspressoituu *C. biroi* -muurahaisella vahvasti tuntosarvissa (McKenzie ym. 2014) ja toimii mehiläisellä tunnetusti kuningatarferomonin sitomisessa (Danty ym. 1999, Weng ym. 2014). OBP1:n funktio liittyy siis todennäköisesti hajumolekyylien sitomiseen myös muurahaisilla. OBP1-geenistä on merkitseviä tuloksia MK ja Fu ja Li -testeistä (Taulukko 26). MK-testin tulos viittaa positiiviseen valintaan kahden lajin, *F. aquilonian* ja *F. lugubriksen*, osalta. Kaukaisen *C. floridanus* -muurahaisen käyttäminen ulkoryhmänä ja siitä johtuva suuri määrä eroja lajien välillä (Liite 1) saattaa kuitenkin vääristää testitulosta adaptiivisen valinnan suuntaan. Näistä lajeista ei ole muiden valinnan testien tuloksia liian pienten populaatio-otoksien vuoksi. Fu ja Li -testin tulos viittaa *F. exsectan* osalta tasapainottavaan valintaan, mutta tulos voi liittyä myös populaation piilorakenteeseen.

OBP10-geenin ekspressioprofiilit muurahaisella (McKenzie ym. 2014) ja mehiläisellä (Forêt ym. 2006) viittaavat siihen, että geenin funktio liittyy johonkin muuhun kuin hajumolekyylien sitomiseen tuntosarvissa. Valinnan testit eivät havaitse OBP10:ssä koko geeniin kohdistuvia valinnan voimia (Taulukko 26). Ainoa merkitsevä tulos (*F. exsectan* Tajiman D -tulos) saattaa johtua populaation piilorakenteesta, eikä valinnasta. Toisaalta monista OBP10-geenin sekvensseistä puuttui osia, mikä pienensi analysoitavaa osaa geenistä ja saattoi vähentää testien tehoa.

MK-, Tajiman D ja Fu ja Li -testeillä on aiemmin havaittu merkkejä valinnasta *Drosophila*-kärpäsen OBP- (Wang ym. 2007) ja OR-geeneissä (Rollmann ym. 2010) ja ihmisen OR-geeneissä (Gilad ym. 2003). Näiden testien tulosten epäjohdonmukaisuus on tullut esille myös *Drosophila*-kärpäsen hajugeenien tutkimuksissa (Wang ym. 2007, Rollmann ym. 2010). Joidenkin OR-geenien evoluutio poikkesi neutraalista mallista MK-testin mukaan, mutta Tajiman D -testin tulokset eivät olleet merkitseviä (Rollmann ym. 2010). OBP-geenien tapauksessa eroavien tulosten pääteltiin johtuvan testien eri aikaskaaloista ja herkkyysistä populaatiodemografialle sekä Tajiman D ja Fu ja Li -testien tehokkuuksien eroista (Wang ym. 2007).

6.2.2. Geenisäätelyyn kohdistuva luonnonvalinta

Luonnonvalinta voi vaikuttaa myös ei-koodaavaan sekvenssiin ja adaptiivinen evoluutio voi tapahtua geenisäätelyn kautta (Brawand ym. 2014, Simola ym. 2013). Esimerkiksi sosiaalisten hyönteisten eri kastit, työläinen ja kuningatar, syntyvät samasta genotyypistä erilaisen geenisäätelyn kautta (Kucharski ym. 2008). Valinta voi vaikuttaa esimerkiksi geenin ekspressiotasoon transkriptiofaktoreiden sitoutumiskohtien kautta (Weingarten-Gabbay ym. 2014). Tajiman D ja Fu ja Li -evoluutiotestien mukaan mahdollisesti valinnan alaisille alueille ennustettiin transkriptiofaktoreiden sitoutumiskohtia. Sitoutumiskohtien määrässä tai laadussa oli jonkin verran muuntelua joko lajien välillä tai sisällä jokaisessa geenissä, mutta vain CSP7:ssä ja OBP10:ssä jokin näistä eroista osui mahdollisen valinnan alueelle. *F. cinerea*lla havaittiin mahdollista positiivista tai puhdistavaa CSP7-geenin intronissa. Tässä kohdassa on lajin sisäistä muuntelua Tll-faktorin sitoutumiskohdassa. *F. cinerea*lla havaittiin mahdollista positiivista tai puhdistavaa myös OBP10-geenin neljännessä intronissa. Tässä kohdassa on lajin sisäistä muuntelua Ftz-faktorin sitoutumiskohdassa. Transkriptiofaktorit Tll eli Tailless ja Ftz eli Fushi tarazu toimivat yksilönkehityksessä *Drosophila*-kärpäsellä ja Ftz toimii lisäksi esimerkiksi aktivaattorina.

6.3. Muurahaisten sosiaalisten muotojen välinen muuntelu

Muurahaisten kaksi sosiaalista muotoa, monogyninen ja polygyninen, eroavat toisistaan monissa käyttäytymispiirteissä (Ross ym. 1995). Näiden muotojen geneettinen tausta on kiinnostanut tutkijoita jo pitkään. Tulimuurahaisen sosiaalikromosomiin kuuluvassa OBP-geenin Gp9 eksoneissa on havaittu 9 SNP:iä, jotka löydetään vain polygynisistä eli usean kuningattaren populaatioista (Krieger ym. 2002). Tämän muuntelun Gp-9-geenissä on ajateltu osaltaan vaikuttavan polygynisten populaatioiden käyttäytymispiirteisiin. Tässä työssä tutkittiin *F. cinerea*n, *F. exsectan* ja *F. truncorum*n osalta, onko lajien sisällä sekvenssimuuntelua nimenomaan monogynisten ja polygynisten muotojen välillä. Sekvensseistä ei löytynyt kohtia, joissa monogynisten ja polygynisten ryhmien välillä olisi fiksoituneita eroja. Sosiaalisten muotojen välillä oli kuitenkin eroa SNP:ien esiintyvyydessä. Aineistossa oli SNP:ejä, jotka olivat läsnä vain polygynisissä näytteissä, mutta eivät monogynisissä näytteissä kuten tulimuurahaisenkin tapauksessa. Kuitenkin vain muutama näistä sijoittui eksoneihin: *F. exsectalla* on kaksi vastaavaa SNP:iä CSP1-geenissä ja yksi OBP1-geenissä. Lisäksi näiden SNP:ien alleelifrekvenssit olivat alhaisia (0,10 – 0,13), joten ne

saattavat esiintyä myös monogynisissä populaatioissa, vaikka niitä ei havaittu tämän työn otoksessa. Tässä aineistossa havaitut erot sosiaalisten muotojen välillä ovat siis hyvin vähäisiä verrattuna tulimuurahaiseen, joten ainakaan tämän aineiston perusteella ei voida olettaa, että mikään geeneistä CSP1, CSP7, OBP1 ja OBP10 kuuluisi vastaavaan sosiaalikromosomiin tutkituilla lajeilla. Myös näitä lajeja läheisemmän *F. selysin* genomista on hiljattain löydetty tulimuurahaisen sosiaalikromosomia vastaava alue, joka on kuitenkin eri evolutiivista alkuperää tulimuurahaisella (Purcell ym. 2014).

6.4. Käytettyjen menetelmien soveltuvuus

Varsinkin OBP10-geenin sekvensseissä on paljon puuttuvia alueita, jotka olivat alun perin lukukelvottomia tai eivät linjautuneet muiden sekvenssifragmenttien kanssa. Esimerkiksi A/T-toistojaksot vaikeuttivat sekvenssien kokoamista. Näytekoiko myös pieneni huomattavasti sekvenssien kasaamisvaiheessa huonolaatuisten sekvenssien vuoksi. Huonolaatuiset sekvenssit voivat johtua esimerkiksi kontaminaatioista tai siitä, etteivät muiden lajien perusteella suunnitellut alukkeet sopineet täysin *Formica*-lajien geeneille. Analysointivaiheessa löytyi myös muutama yksittäinen näyte, jotka erosivat kaikista muista saman lajin näytteistä ja muistuttivat jonkin muun lajin näytteitä eli todennäköisesti oli tapahtunut lajinmäärittämisvirhe. Tällaiset näytteet poistettiin. Analyysivaiheessa sekvensseissä ei ollut mukana tietoa niiden laadusta. Löydetty SNP:it voivat siis olla epävarmoja. Sellaiset SNP:it, joita löydetään useista yksilöistä eli jotka esiintyvät korkealla alleelifrekvenssillä, ovat suurella varmuudella kuitenkin todellisia. Myös kytkentäfaasin (linkage phase) arvioiminen tilastollisin menetelmin on epävarmaa. Monet analyysit eivät huomioi indeleitä eli insertioita ja deleetioita.

Luonnonvalinnan testien tulokset eivät ole yhdenmukaisia. Yksittäisten lajien osalta osa testituloksista on merkitseviä, osa ei ja osa merkitsevistä tuloksista on ristiriidassa keskenään. Pienen aineiston ja vähäisen muuntelun vuoksi tällaiset tulokset ovat odotettuja. Jatkossa tulisikin varmistaa, että tutkittavasta lajista voidaan kerätä suurempi näyte. Lisäksi eri testit käyttävät geenin eri osia, tunnistavat valintaa eri aikaskaaloissa, tunnistavat erityyppistä valintaa ja eroavat herkkyydessä populaatorakenteelle ja ulkoryhmän käyttämisessä. MK-testissä on mukana vain geenien eksonit. Muissa evoluutioanalyysissä on käytetty koko geenejä eli niissä on myös intronit mukana. MK-testi tunnistaa aiempaa, lajien välillä tapahtunutta valintaa ja muut testit taas viimeaikaista valintaa (Vitti ym. 2013, Li 2011). Tajiman D -testi ei käytä ulkoryhmää toisin kuin

muut testit. Lisäksi käytetty ulkoryhmä vaihteli testien välillä. MK-testissä ja Fu ja Li -testissä käytettiin ulkoryhmänä *Camponotus floridanus* ja MFDm-testissä käytettiin ulkoryhmänä jotain toista aineiston lajeista. *Camponotus floridanus* saattaa olla liian kaukainen tutkituista lajeista ollakseen sopiva ulkoryhmä.

Tajiman D ja Fu ja Li -testit ovat muita testejä herkempiä populaatiokoon muutoksille ja populaatorakenteelle (Simonsen ym. 1995). *F. exsectan* ja *F. truncorumin* monogyniset populaatiot yhdistettiin suuremman aineiston aikaansaamiseksi. Näiden lajien CSP7, OBP1- ja OBP10-geenien merkitsevät Tajiman D ja Fu ja Li -testien tulokset voivat johtua myös populaatorakenteesta. Populaation piilorakenne voi tuottaa ylimäärän harvinaisia variantteja, joita testit detektoivat. Monogynisten populaatioiden yhdistäminen ei kuitenkaan vaikuta muihin testeihin (MK ja MFDm), eikä muihin lajeihin (*F. cinerean* populaatioita ei yhdistetty).

Sekvenssin eri osat voivat olla eri valinnan voimien vaikutuksen alaisina. Esimerkiksi eksonit voivat olla puhdistavan valinnan alaisena ja intronit positiivisen valinnan alaisena. Tällöin muutoksia geenin toiminnassa eivät ajaisikaan proteiinin rakenteen muutokset, vaan geenisäätelyn muutokset. Tällöin sliding window -tyyppiset testit voivat antaa realistisempia tuloksia, koska koko geenin ei odotetakaan olevan samantyyppisten evoluutiovoimien alaisena.

Transkriptiofaktoreiden sitoutumiskohtia voitiin ennustaa vain introneista, koska promoottorialueiden sekvenssejä ei ollut käytettävissä. Sitoutumiskohtien laskennallinen ennustaminen on epävarmaa. Transkriptiofaktorit eivät välttämättä sitoudu ennustettuihin kohtiin. Tulokset pitäisi siis tarkistaa kokeellisesti ennen kuin niistä voi tehdä johtopäätöksiä. Toisaalta sitoutumiskohtien ennustamisessa vaadittiin sekvenssin täyttä identtisuutta, joten todellisia sitoutumiskohtia voi olla myös enemmän.

6.5. Johtopäätökset ja jatkotutkimukset

Tutkimuksen seitsemässä läheisessä *Formica*-muurahaislajissa on vain vähän muuntelua tutkituissa CSP- ja OBP-geeneissä. *F. cinerea* ja *F. exsecta* eroavat muista lajeista ja toisistaan, mutta *F. rufa* -ryhmän lajeista ainoastaan *F. truncorum* erottuu omaksi ryhmäkseen. Tämä voi johtua viimeaikaisesta lajiutumisesta tai lajien välisestä risteytymisestä. *F. rufa* -ryhmän sekvenssien samankaltaisuuden syytä voisi tutkia jatkossa vertailemalla vähemmän

konservoituneita geenejä tai ei-koodaavia sekvenssejä. Jos niissä löydetään fiksoituneita eroja *F. rufa* -ryhmän lajien välillä, lajien samankaltaisuus hajugeeneissä johtuu todennäköisesti hiljattaisesta lajiutumisesta. Toisaalta jos lajit risteytyvät vielä keskenään, muistakaan geeneistä/sekvensseistä ei pitäisi löytyä paljon eroja.

Pesätoverien tunnistuksessa toimivassa CSP7-geenissä on eniten merkkejä sekä koko geeniin että tiettyihin alueisiin vaikuttavista valinnan voimista. Näistä *F. cinerean* tulokset ovat luotettavimpia suurimman aineiston ja populaation piilorakenteen puutteen vuoksi. Varsinkin *F. cinerean* CSP7:n introniin sijoittuva mahdollinen, valinnan alainen transkriptiofaktorin sitoutumiskohta on kiinnostava jatkotutkimuksia ajatellen. Sitoutuuko tähän kohtaan todella transkriptiofaktori ja vaikuttaako *F. cinerean* SNP sen sitoutumiseen ja CSP7:n ekspressioon? Valinnan alaisuudessa kehittyville alueille ennustetut transkriptiofaktoreiden sitoutumiskohdat kannattaisi siis varmistaa kokeellisesti. Transkriptiofaktoreiden sitoutumiskohtien ennustaminen kannattaisi tehdä myös geenien promoottorialueille.

7. KIITOKSET

Kiitos työn rahoituksesta Vanamo ry:lle ja Societas pro Fauna et Flora Fennica:lle. Tutkimus tehtiin Helsingin yliopistossa, Biotieteiden laitoksella, Antzz-ryhmässä. Sekvenssiaineisto tuotettiin Oulun yliopistossa professori Pekka Pamilon ryhmässä. Kiitos näytteistä Pekka Pamilolle, Anu Sirviölle, David Hughesille ja Anya Goropashnayalle ja laboratoriotyöstä Riitta Jokelalle. Kiitos työn ohjauksesta ja neuvoista tohtori Jonna Kulmunille.

8. LÄHTEET

- Andolfatto, P. 2005, "Adaptive evolution of non-coding DNA in *Drosophila*", *Nature*, vol. 437, no. 7062, pp. 1149-1152.
- Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I. & Stockinger, H. 2012, "ExPASy: SIB bioinformatics resource portal", *Nucleic acids research*, vol. 40, no. Web Server issue, pp. W597-603.
- Birney, E., Clamp, M. & Durbin, R. 2004, "GeneWise and Genomewise", *Genome research*, vol. 14, no. 5, pp. 988-995.
- Bonasio, R., Zhang, G., Ye, C., Mutti, N.S., Fang, X., Qin, N., Donahue, G., Yang, P., Li, Q., Li, C., Zhang, P., Huang, Z., Berger, S.L., Reinberg, D., Wang, J. & Liebig, J. 2010, "Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*", *Science (New York, N.Y.)*, vol. 329, no. 5995, pp. 1068-1071.
- Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H.J., Russell, P., Aken, B., Alföldi, J., Amemiya, C., Azzouzi, N., Baroiller, J.F., Barloy-Hubler, F., Berlin, A., Bloomquist, R., Carleton, K.L., Conte, M.A., D'Cotta, H., Eshel, O., Gaffney, L., Galibert, F., Gante, H.F., Gnerre, S., Greuter, L., Guyon, R., Haddad, N.S., Haerty, W., Harris, R.M., Hofmann, H.A., Hourlier, T., Hulata, G., Jaffe, D.B., Lara, M., Lee, A.P., MacCallum, I., Mwaiko, S., Nikaido, M., Nishihara, H., Ozouf-Costaz, C., Penman, D.J., Przybylski, D., Rakotomanga, M., Renn, S.C., Ribeiro, F.J., Ron, M., Salzburger, W., Sanchez-Pulido, L., Santos, M.E., Searle, S., Sharpe, T., Swofford, R., Tan, F.J., Williams, L., Young, S., Yin, S., Okada, N., Kocher, T.D., Miska, E.A., Lander, E.S., Venkatesh, B., Fernald, R.D., Meyer, A., Ponting, C.P., Streelman, J.T., Lindblad-Toh, K., Seehausen, O. & Di Palma, F. 2014, "The genomic substrate for adaptive radiation in African cichlid fish", *Nature*, vol. 513, no. 7518, pp. 375-381.
- Brown, W.M., George, M., Jr & Wilson, A.C. 1979, "Rapid evolution of animal mitochondrial DNA", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 76, no. 4, pp. 1967-1971.
- Bustamante, C.D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M.T., Glanowski, S., Tanenbaum, D.M., White, T.J., Sninsky, J.J., Hernandez, R.D., Civello, D., Adams, M.D., Cargill, M. & Clark, A.G. 2005, "Natural selection on protein-coding genes in the human genome", *Nature*, vol. 437, no. 7062, pp. 1153-1157.
- Carneiro, M., Albert, F.W., Melo-Ferreira, J., Galtier, N., Gayral, P., Blanco-Aguilar, J.A., Villafuerte, R., Nachman, M.W. & Ferrand, N. 2012, "Evidence for widespread positive and purifying selection across the European rabbit (*Oryctolagus cuniculus*) genome", *Molecular biology and evolution*, vol. 29, no. 7, pp. 1837-1849.
- Charlesworth, B., Bartolome, C. & Noel, V. 2005, "The detection of shared and ancestral polymorphisms", *Genetical research*, vol. 86, no. 2, pp. 149-157.

- Danty, E., Briand, L., Michard-Vanhee, C., Perez, V., Arnold, G., Gaudemer, O., Huet, D., Huet, J.C., Ouali, C., Masson, C. & Pernollet, J.C. 1999, "Cloning and expression of a queen pheromone-binding protein in the honeybee: an olfactory-specific, developmentally regulated protein", *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 19, no. 17, pp. 7468-7475.
- Delport, W., Poon, A.F., Frost, S.D. & Kosakovsky Pond, S.L. 2010, "Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology", *Bioinformatics (Oxford, England)*, vol. 26, no. 19, pp. 2455-2457.
- Eilertson, K.E., Booth, J.G. & Bustamante, C.D. 2012, "SnIPRE: selection inference using a Poisson random effects model", *PLoS computational biology*, vol. 8, no. 12, pp. e1002806.
- Excoffier, L. & Lischer, H.E. 2010, "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows", *Molecular ecology resources*, vol. 10, no. 3, pp. 564-567.
- Farré, D., Roset, R., Huerta, M., Adsuara, J.E., Rosello, L., Alba, M.M. & Messeguer, X. 2003, "Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN", *Nucleic acids research*, vol. 31, no. 13, pp. 3651-3653.
- Ferreira, M.U. & Hartl, D.L. 2007, "Plasmodium falciparum: worldwide sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-2 (MSP-2)", *Experimental parasitology*, vol. 115, no. 1, pp. 32-40.
- Flot, J.F. 2010, "Seqphase: a Web Tool for Interconverting Phase Input/output Files and Fasta Sequence Alignments", *Molecular ecology resources*, vol. 10, no. 1, pp. 162-166.
- Forêt, S. & Maleszka, R. 2006, "Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*)", *Genome research*, vol. 16, no. 11, pp. 1404-1413.
- Forêt, S., Wanner, K.W. & Maleszka, R. 2007, "Chemosensory proteins in the honey bee: Insights from the annotated genome, comparative analyses and expressional profiling", *Insect biochemistry and molecular biology*, vol. 37, no. 1, pp. 19-28.
- Fu, Y.X. 1997, "Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection", *Genetics*, vol. 147, no. 2, pp. 915-925.
- Fu, Y.X. & Li, W.H. 1993, "Statistical tests of neutrality of mutations", *Genetics*, vol. 133, no. 3, pp. 693-709.
- Gilad, Y., Bustamante, C.D., Lancet, D. & Paabo, S. 2003, "Natural selection on the olfactory receptor gene family in humans and chimpanzees", *American Journal of Human Genetics*, vol. 73, no. 3, pp. 489-501.
- Goropashnaya, A.V., Fedorov, V.B. & Pamilo, P. 2004, "Recent speciation in the *Formica rufa* group ants (Hymenoptera, Formicidae): inference from mitochondrial DNA phylogeny", *Molecular phylogenetics and evolution*, vol. 32, no. 1, pp. 198-206.

- Goropashnaya, A.V., Fedorov, V.B., Seifert, B. & Pamilo, P. 2012, "Phylogenetic relationships of Palaearctic *Formica* species (Hymenoptera, Formicidae) based on mitochondrial cytochrome B sequences", *PLoS one*, vol. 7, no. 7, pp. e41697.
- Haddrill, P.R., Bachtrog, D. & Andolfatto, P. 2008, "Positive and negative selection on noncoding DNA in *Drosophila simulans*", *Molecular biology and evolution*, vol. 25, no. 9, pp. 1825-1834.
- Hall, T.A. 1999, "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT", *Nucleic Acids Symposium Series*, vol. 41, pp. 95-98.
- Harpur, B.A., Kent, C.F., Molodtsova, D., Lebon, J.M., Alqarni, A.S., Owayss, A.A. & Zayed, A. 2014, "Population genomics of the honey bee reveals strong signatures of positive selection on worker traits", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 7, pp. 2614-2619.
- Hartl, D.L. & Clark, A.G. 2007, *Principles of population genetics*, 4th edition. edn, Sinauer Associates.
- Katoh, K. & Standley, D.M. 2013, "MAFFT multiple sequence alignment software version 7: improvements in performance and usability", *Molecular biology and evolution*, vol. 30, no. 4, pp. 772-780.
- Kimura, M. 1968, "Evolutionary rate at the molecular level", *Nature*, vol. 217, no. 5129, pp. 624-626.
- Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H. & Frost, S.D. 2006, "Automated phylogenetic detection of recombination using a genetic algorithm", *Molecular biology and evolution*, vol. 23, no. 10, pp. 1891-1901.
- Krieger, M.J. & Ross, K.G. 2005, "Molecular evolutionary analyses of the odorant-binding protein gene Gp-9 in fire ants and other Solenopsis species", *Molecular biology and evolution*, vol. 22, no. 10, pp. 2090-2103.
- Krieger, M.J. & Ross, K.G. 2002, "Identification of a major gene regulating complex social behavior", *Science (New York, N.Y.)*, vol. 295, no. 5553, pp. 328-332.
- Kucharski, R., Maleszka, J., Foret, S. & Maleszka, R. 2008, "Nutritional control of reproductive status in honeybees via DNA methylation", *Science (New York, N.Y.)*, vol. 319, no. 5871, pp. 1827-1830.
- Kulmuni, J. & Havukainen, H. 2013b, "Insights into the evolution of the CSP gene family through the integration of evolutionary analysis and comparative protein modeling", *PLoS one*, vol. 8, no. 5, pp. e63688.
- Kulmuni, J., Seifert, B. & Pamilo, P. 2010, "Segregation distortion causes large-scale differences between male and female genomes in hybrid ants", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 16, pp. 7371-7376.
- Kulmuni, J., Wurm, Y. & Pamilo, P. 2013a, "Comparative genomics of chemosensory protein genes reveals rapid evolution and positive selection in ant-specific duplicates", *Heredity*, vol. 110, no. 6, pp. 538-547.

- Laughlin, J.D., Ha, T.S., Jones, D.N. & Smith, D.P. 2008, "Activation of pheromone-sensitive neurons is mediated by conformational activation of pheromone-binding protein", *Cell*, vol. 133, no. 7, pp. 1255-1265.
- Leal, W.S. 2013, "Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes", *Annual Review of Entomology*, vol. 58, pp. 373-391.
- Leal, W.S. & Ishida, Y. 2008, "GP-9s are ubiquitous proteins unlikely involved in olfactory mediation of social organization in the red imported fire ant, *Solenopsis invicta*", *PloS one*, vol. 3, no. 11, pp. e3762.
- Li, H. 2011, "A new test for detecting recent positive selection that is free from the confounding impacts of demography", *Molecular biology and evolution*, vol. 28, no. 1, pp. 365-375.
- Li, J., Li, H., Jakobsson, M., Li, S., Sjödin, P. & Lascoux, M. 2012, "Joint analysis of demography and selection in population genetics: where do we stand and where could we go?", *Molecular ecology*, vol. 21, no. 1, pp. 28-44.
- Librado, P. & Rozas, J. 2009, "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data", *Bioinformatics (Oxford, England)*, vol. 25, no. 11, pp. 1451-1452.
- Löbel, D., Jacob, M., Volkner, M. & Breer, H. 2002, "Odorants of different chemical classes interact with distinct odorant binding protein subtypes", *Chemical senses*, vol. 27, no. 1, pp. 39-44.
- McDonald, J.H. & Kreitman, M. 1991, "Adaptive protein evolution at the Adh locus in *Drosophila*", *Nature*, vol. 351, no. 6328, pp. 652-654.
- McKenzie, S.K., Oxley, P.R. & Kronauer, D.J. 2014, "Comparative genomics and transcriptomics in ants provide new insights into the evolution and function of odorant binding and chemosensory proteins", *BMC genomics*, vol. 15, pp. 718-2164-15-718.
- Munoz-Torres, M.C., Reese, J.T., Childers, C.P., Bennett, A.K., Sundaram, J.P., Childs, K.L., Anzola, J.M., Milshina, N. & Elsik, C.G. 2011, "Hymenoptera Genome Database: integrated community resources for insect species of the order Hymenoptera", *Nucleic acids research*, vol. 39, no. Database issue, pp. D658-62.
- Nei, M. 2005, "Selectionism and neutralism in molecular evolution", *Molecular biology and evolution*, vol. 22, no. 12, pp. 2318-2342.
- Ohta, T. 1992, "The nearly Neutral Theory of Molecular Evolution", *Annual Review of Ecology and Systematics*, vol. 23, pp. 263-286.
- Ohta, T. 1973, "Slightly deleterious mutant substitutions in evolution", *Nature*, vol. 246, no. 5428, pp. 96-98.
- Ozaki, M., Wada-Katsumata, A., Fujikawa, K., Iwasaki, M., Yokohari, F., Satoji, Y., Nisimura, T. & Yamaoka, R. 2005, "Ant nestmate and non-nestmate discrimination by a chemosensory sensillum", *Science (New York, N.Y.)*, vol. 309, no. 5732, pp. 311-314.

- Peakall, R. & Smouse, P.E. 2012, "GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update", *Bioinformatics (Oxford, England)*, vol. 28, no. 19, pp. 2537-2539.
- Peakall, R. & Smouse, P.E. 2006, "genalex 6: genetic analysis in Excel. Population genetic software for teaching and research", *Molecular Ecology Notes*, vol. 6, no. 1, pp. 288-295.
- Pelosi, P., Calvello, M. & Ban, L. 2005, "Diversity of odorant-binding proteins and chemosensory proteins in insects", *Chemical senses*, vol. 30 Suppl 1, pp. i291-2.
- Pond, S.L. & Frost, S.D. 2005b, "Datamonkey: rapid detection of selective pressure on individual sites of codon alignments", *Bioinformatics (Oxford, England)*, vol. 21, no. 10, pp. 2531-2533.
- Pond, S.L., Frost, S.D. & Muse, S.V. 2005a, "HyPhy: hypothesis testing using phylogenies", *Bioinformatics (Oxford, England)*, vol. 21, no. 5, pp. 676-679.
- Prasad, B.C. & Reed, R.R. 1999, "Chemosensation: molecular mechanisms in worms and mammals", *Trends in genetics : TIG*, vol. 15, no. 4, pp. 150-153.
- Purcell, J., Brelsford, A., Wurm, Y., Perrin, N. & Chapuisat, M. 2014, "Convergent genetic architecture underlies social organization in ants", *Current biology : CB*, vol. 24, no. 22, pp. 2728-2732.
- Rand, D.M. & Kann, L.M. 1996, "Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans", *Molecular biology and evolution*, vol. 13, no. 6, pp. 735-748.
- Richard, F.-. & Hunt, J.H. 2013, "Intracolony chemical communication in social insects", *Insectes Sociaux*, vol. 60, no. 3, pp. 275-291.
- Rollmann, S.M., Wang, P., Date, P., West, S.A., Mackay, T.F. & Anholt, R.R. 2010, "Odorant receptor polymorphisms and natural variation in olfactory behavior in *Drosophila melanogaster*", *Genetics*, vol. 186, no. 2, pp. 687-697.
- Ross, K. & Keller, L. 1995, "Ecology and Evolution of Social-Organization - Insights from Fire Ants and Other Highly Eusocial Insects", *Annual Review of Ecology and Systematics*, vol. 26, pp. 631-656.
- Salazar-Jaramillo, L., Paspatis, A., van de Zande, L., Vermeulen, C.J., Schwander, T. & Wertheim, B. 2014, "Evolution of a cellular immune response in *Drosophila*: a phenotypic and genomic comparative analysis", *Genome biology and evolution*, vol. 6, no. 2, pp. 273-289.
- Simola, D.F., Wissler, L., Donahue, G., Waterhouse, R.M., Helmkampf, M., Roux, J., Nygaard, S., Glastad, K.M., Hagen, D.E., Viljakainen, L., Reese, J.T., Hunt, B.G., Graur, D., Elhaik, E., Kriventseva, E.V., Wen, J., Parker, B.J., Cash, E., Privman, E., Childers, C.P., Munoz-Torres, M.C., Boomsma, J.J., Bornberg-Bauer, E., Currie, C.R., Elsik, C.G., Suen, G., Goodisman, M.A., Keller, L., Liebig, J., Rawls, A., Reinberg, D., Smith, C.D., Smith, C.R., Tsutsui, N., Wurm, Y., Zdobnov, E.M., Berger, S.L. & Gadau, J. 2013, "Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality", *Genome research*, vol. 23, no. 8, pp. 1235-1247.

- Simonsen, K.L., Churchill, G.A. & Aquadro, C.F. 1995, "Properties of statistical tests of neutrality for DNA polymorphism data", *Genetics*, vol. 141, no. 1, pp. 413-429.
- Singer, T.L. 1998, "Roles of Hydrocarbons in the Recognition Systems of Insects", *American Zoologist*, vol. 38, no. 2, pp. 394-405.
- Sirviö, A. & Pamilo, P. 2010, "Multiple endosymbionts in populations of the ant *Formica cinerea*", *BMC evolutionary biology*, vol. 10, pp. 335-2148-10-335.
- Smadja, C. & Butlin, R.K. 2009, "On the scent of speciation: the chemosensory system and its role in premating isolation", *Heredity*, vol. 102, no. 1, pp. 77-97.
- Smith, N.G. & Eyre-Walker, A. 2002, "Adaptive protein evolution in *Drosophila*", *Nature*, vol. 415, no. 6875, pp. 1022-1024.
- Stephens, M. & Scheet, P. 2005, "Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation", *American Journal of Human Genetics*, vol. 76, no. 3, pp. 449-462.
- Stephens, M., Smith, N.J. & Donnelly, P. 2001, "A new statistical method for haplotype reconstruction from population data", *American Journal of Human Genetics*, vol. 68, no. 4, pp. 978-989.
- Swarup, S., Williams, T.I. & Anholt, R.R. 2011, "Functional dissection of Odorant binding protein genes in *Drosophila melanogaster*", *Genes, brain, and behavior*, vol. 10, no. 6, pp. 648-657.
- Symonds, M.R. & Elgar, M.A. 2008, "The evolution of pheromone diversity", *Trends in ecology & evolution*, vol. 23, no. 4, pp. 220-228.
- Tajima, F. 1989, "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism", *Genetics*, vol. 123, no. 3, pp. 585-595.
- Tegoni, M., Pelosi, P., Vincent, F., Spinelli, S., Campanacci, V., Grolli, S., Ramoni, R. & Cambillau, C. 2000, "Mammalian odorant binding proteins", *Biochimica et biophysica acta*, vol. 1482, no. 1-2, pp. 229-240.
- Vieira, F.G., Forêt, S., He, X., Rozas, J., Field, L.M. & Zhou, J. 2012, "Unique Features of Odorant-Binding Proteins of the Parasitoid Wasp *Nasonia vitripennis* Revealed by Genome Annotation and Comparative Analyses", *PLoS ONE*, vol. 7, no. 8, pp. e43034.
- Vitti, J.J., Grossman, S.R. & Sabeti, P.C. 2013, "Detecting natural selection in genomic data", *Annual Review of Genetics*, vol. 47, pp. 97-120.
- Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.C., Shoemaker, D. & Keller, L. 2013, "A Y-like social chromosome causes alternative colony organization in fire ants", *Nature*, vol. 493, no. 7434, pp. 664-668.
- Wang, P., Lyman, R.F., Shabalina, S.A., Mackay, T.F. & Anholt, R.R. 2007, "Association of polymorphisms in odorant-binding protein genes with variation in olfactory response to benzaldehyde in *Drosophila*", *Genetics*, vol. 177, no. 3, pp. 1655-1665.

- Wasserman, W.W. & Sandelin, A. 2004, "Applied bioinformatics for the identification of regulatory elements", *Nature reviews.Genetics*, vol. 5, no. 4, pp. 276-287.
- Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J. 2009, "Jalview Version 2--a multiple sequence alignment editor and analysis workbench", *Bioinformatics (Oxford, England)*, vol. 25, no. 9, pp. 1189-1191.
- Weingarten-Gabbay, S. & Segal, E. 2014, "The grammar of transcriptional regulation", *Human genetics*, vol. 133, no. 6, pp. 701-711.
- Weng, C., Fu, Y., Jiang, H., Zhuang, S. & Li, H. 2014, "Binding interaction between a queen pheromone component HOB and pheromone binding protein ASP1 of *Apis cerana*", *International journal of biological macromolecules*, vol. 72C, pp. 430-436.
- Whiteman, N.K. & Pierce, N.E. 2008, "Delicious poison: genetics of *Drosophila* host plant preference", *Trends in ecology & evolution*, vol. 23, no. 9, pp. 473-478.
- Wurm, Y., Wang, J., Riba-Grognuz, O., Corona, M., Nygaard, S., Hunt, B.G., Ingram, K.K., Falquet, L., Nipitwattanaphon, M., Gotzek, D., Dijkstra, M.B., Oettler, J., Comtesse, F., Shih, C.J., Wu, W.J., Yang, C.C., Thomas, J., Beaudoin, E., Pradervand, S., Flegel, V., Cook, E.D., Fabbretti, R., Stockinger, H., Long, L., Farmerie, W.G., Oakey, J., Boomsma, J.J., Pamilo, P., Yi, S.V., Heinze, J., Goodisman, M.A., Farinelli, L., Harshman, K., Hulo, N., Cerutti, L., Xenarios, I., Shoemaker, D. & Keller, L. 2011, "The genome of the fire ant *Solenopsis invicta*", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 14, pp. 5679-5684.
- Xu, W. & Leal, W.S. 2008, "Molecular switches for pheromone release from a moth pheromone-binding protein", *Biochemical and biophysical research communications*, vol. 372, no. 4, pp. 559-564.

Ohjelmien Internet-osoitteet:

Ant Genome Portal: http://hymenopteragenome.org/ant_genomes

CodonCode Aligner: <http://www.codoncode.com>

MAFFT: <http://mafft.cbrc.jp/alignment/software>

Jalview: <http://www.jalview.org>

Bioedit: <http://www.mbio.ncsu.edu/bioedit/bioedit.html>

Genewise: <http://www.ebi.ac.uk/Tools/psa/genewise>

ExPASy Translate: <http://web.expasy.org/translate>

seqPhase: <http://seqphase.mpg.de/seqphase>

MEGA: <http://www.megasoftware.net>

FigTree: <http://tree.bio.ed.ac.uk/software/figtree>

HyPhy: <http://www.datamonkey.org>

ALGGEN PROMO: <http://alggen.lsi.upc.es>

9. LIITTEET

Liite 1. MK-testitulostaulukko.

Geeni	Laji	Syn. fiksoituneet erot	Ei-syn. fiks. erot	Syn. polymorfismit	Ei-syn. polymorf.
CSP1	<i>F. aquilonia</i>	24	8	1	0
	<i>F. cinerea</i>	21	6	3	2
	<i>F. exsecta</i>	22	8	2	0
	<i>F. lugubris</i>	24	8	1	0
	<i>F. polycltena</i>	23	8	2	1
	<i>F. pratensis</i>	24	8	0	2
	<i>F. rufa</i>	24	8	1	0
	<i>F. truncorum</i>	23	8	1	1
CSP7	<i>F. aquilonia</i>	26	11	0	0
	<i>F. cinerea</i>	23	10	3	8
	<i>F. exsecta</i>	27	11	0	0
	<i>F. lugubris</i>	26	11	0	0
	<i>F. polycltena</i>	24	10	3	0
	<i>F. rufa</i>	26	11	3	1
	<i>F. truncorum</i>	22	11	1	0
OBP1	<i>F. aquilonia</i>	22	30	4	0
	<i>F. cinerea</i>	22	30	1	1
	<i>F. exsecta</i>	23	30	3	1
	<i>F. lugubris</i>	22	30	5	1
	<i>F. polycltena</i>	22	29	3	0
	<i>F. rufa</i>	18	28	2	0
	<i>F. truncorum</i>	22	31	0	0
OBP10	<i>F. aquilonia</i>	19	7	2	3
	<i>F. cinerea</i>	19	6	2	3
	<i>F. exsecta</i>	18	6	3	0
	<i>F. lugubris</i>	20	15	2	3
	<i>F. polycltena</i>	15	5	2	3
	<i>F. rufa</i>	21	14	1	0
	<i>F. truncorum</i>	19	7	1	2